

TITLE
JOURNAL
COMMENT

CP	1204	GACAGTAATCAATTTTATTTTGTGTCTACAGAACTACTAGGCGATCTCGAGAGTCGCTC	1145
Db	64	CGTGACAGCCCAACCAACCCCAACCCCTTACTCTGCGACCCACCCCTTAAAGGCGACTTCAAG	123
CP	1144	CGTGACACCCCAACCAACCCCAACCCCTTACTCTGCGACCCACCCCTTAAAGGCGACTTCAAG	1085
Db	124	AAGATGGAAGGATCTCAGAGGATCTCATTTCTTAATGTCGCCGGAAGTCTCACACAGTAGA	183
CP	1084	AAGATGGAAGGATCTCAGAGGATCTCATTTCTTAATGTCGCCGGAAGTCTCACACAGTAGA	1025
Db	184	CAGACGAGTGTGAGATGCTGGAGAGATGCGACATCCCTTAACCTTACGACCCACCAACG	243
CP	1024	CAGACGAGTGTGAGATGCTGGAGAGATGCGACATCCCTTAACCTTACGACCCACCAACG	965
Db	244	ACTTCATCCAGCCGCGGAGTCCTCCCGACCCGAGTCCTCCCATTTCTTCTCTACTT	303
CP	964	ACTTCATCCAGCCGCGGAGTCCTCCCGACCCGAGTCCTCCCATTTCTTCTCTACTT	905
Db	304	TGCGCGAATTCAGGTGTCTCTGCTTCCACACAGTCCCAACAAAGCTTAATAAT	355
CP	904	TGCGCGAATTCAGGTGTCTCTGCTTCCACACAGTCCCAACAAAGCTTAATAAT	853
RESULT	14		
LOCUS	T49532	405 bp	mRNA EST 08-FEB-1995
DEFINITION	ya/6112.s1 Homo sapiens cDNA clone 67631 3'.		
ACCESSION	T49532		
NID	9651392		
KEYWORDS	EST.		
SOURCE	human clone=67631 library=Stratagene placenta (#937225) vector=pluescript SK- host=SOB cells (kanamycin resistant) primer=-21m13 Ralte1-Ecor1-Ralte2-Xho1 Placental tissue from a Caucasian male. Cloned unidirectionally. Primer: Oligo dt. Average insert size: 1.2 kb; Uni-ZAP XR Vector; 5' adaptor sequence: 5'-GAATTGGGCGACGAG-3'; 3' adaptor sequence: 5'-CTCGAGCTTTTCTTTTCTTTT-3'.		
ORGANISM	Homo sapiens		
REFERENCE	Eucaryote; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 405) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,J., Hultman,M., Kuchaba,T., Le,M., Lennon,G., Maitra,M., Parsons,J., Riklin,L., Roehling,T., Tan,F., Trevisks,E., Waterson,R., Williamson,A., Wohldmann,P. and Wilson,R.		
TITLE	Mashu-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Other-ESTs: ya76f12.r1 Contact: Wilson RK Mashu-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Location/Qualifiers 1..405 /organism="Homo sapiens" /clone="67631"		
FEATURES	Source		
BASE COUNT	108 a 132 c 81 g 82 t 2 others		
ORIGIN			
Query Match	28.1%; Score 345; DB 9; Length 405;		
Best Local Similarity	96.3%; Pred. No. 0.00e+00;		
Matches	388; Conservative 9; Mismatches 9; Indels 6; Gaps 6;		
Db	1	TGTGTTCACAAATACTAGGCGATCTGAGACAGTCGTCGCTGACAGCCCAACCCG	60
CP	1184	TGTGTTCACAAATACTAGGCGATCTGAGACAGTCGTCGCTGACAGCCCAACCCG	1125

Db 61 CAACCTTACCTGCGACGACCAACCTTAAGGCGACTTCAAGAGATGAGAGATCTCACAG 120
 Cp 1124 CAACCTTACCTGCGACGACCAACCTTAAGGCGACTTCAAGAGATGAGAGATCTCACAG 1065
 Db 121 ATCTCATCTTAATGGTCCGCGAAGTCTCACAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 Cp 1064 ATCTCATCTTAATGGTCCGCGAAGTCTCACAGAGAGAGAGAGAGAGAGAGAGAGAG 1005
 Db 181 GAGATGAG 240
 Cp 1004 GAGATGAG 945
 Db 241 TCCTCCCGCCCGAG 300
 Cp 944 TCCTCCCGCCCGAG 886
 Db 301 CTGCTTCACGAGTCCGCGAAGTCTTAATTAATTAATTAATTAATTAATTAATTAAT 360
 Cp 885 CTGCTTCACGAGTCCGCGAAGTCTTAATTAATTAATTAATTAATTAATTAATTAAT 829
 Db 361 GAGGGAACATTTNACACCTTGGCATAGGTTTAAATTAAT 403
 828 GAGGGAACATTTNACACCTTGGCATAGGTTTAAATTAAT 788

RESULT 15
 LOCUS AA301628 413 bp mRNA EST 18-APR-1997
 DEFINITION EST14813 Aorta endothelial cells Homo sapiens cDNA 5' end.
 ACCESSION AA301628
 MID g1954112
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Euthera; Primates; Catarrhini; Homidae;
 Homo.

REFERENCE 1 (bases 1 to 413)
 AUTHORS Adams,M.D., Kerlavage,A.R., Fieschmann,R.D., Fuldner,R.A.,
 Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gockayne,J.D.,
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
 Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
 Fitzgerald,L.M., Fitzhugh,W.M., Fitchman,J.L., Geoghagan,N.S.,
 Glodet,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
 Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
 Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R.,
 Small,K.V., Spriggs,T.A., Utechtback,T.R., Weidman,J.F., Li,Y.,
 Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
 Dake,D., Feng,D.-F., Ferlie,A., Fischer,C., Hastings,G.A.,
 He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
 Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
 Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
 Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
 Fraser,C.M., and Venter,J.C.
 TITLE Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
 MEDLINE 96026280
 COMMENT Other ESTs: TH0175266
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org

FEATURES
 1. 413
 Location/Qualifiers
 source
 /organism="Homo sapiens"

/note="Organ: aorta; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI"
 /db_xref="ATCC (inhost):113941"
 /db_xref="taxon:9606"
 /clone_lib="Aorta endothelial cells"
 /cell_type="endothelial cell"
 /dev_stage="adult"
 mRNA
 BASE COUNT 105 a 108 c 99 g 95 t 6 others
 ORIGIN
 Query Match 27.0%; Score 332; DB 18; Length 413;
 Best Local Similarity 96.8%; Pred. No. 0.00e+00;
 Matches 364; Conservative 0; Mismatches 7; Indels 5; Gaps 5;
 Db 1 CTGACCTGACGACTTCTCCGACCAACACCGGCGCCTTGACTGTAAGTCTGAGAGAG 60
 Qy 427 CTGACCTGACGACTTCTCCGACCAACACCGGCGCCTTGACTGTAAGTCTGAGAGAG 486
 Db 61 CTGACCAAAAGATCGGAAACCAAGAAACCAAGAAACCAAGAAACCAAGAAACCAAG 120
 Qy 487 CTGACCAAAAGATCGGAAACCAAGAAACCAAGAAACCAAGAAACCAAGAAACCAAG 546
 Db 121 N-TATCCCACTGGAGACTCCGAGCAACTTGAAGTCTGAGAGAGAGAGAGAGAGAG 179
 Qy 547 CGTATCCCACTGGAGACTCCGAGCAACTTGAAGTCTGAGAGAGAGAGAGAGAGAGAG 606
 Db 180 ACCGGTCTGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 239
 Qy 607 ACCGGTCTGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 666
 Db 240 CAGCCGACGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 299
 Qy 667 CAGCCGACGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 725
 Db 300 CTNCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 359
 Qy 726 CTCTT-AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 782
 Db 360 ACCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 375
 Qy 783 ACCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 798

Search completed: Thu Apr 23 05:39:21 1998
 Job time: 811 secs.

Query Match 8.8%; Score 98; DB 1; Length 475;
 Best Local Similarity 28.6%; Pred. No. 6.05e-01;
 Matches 18; Conservative 19; Mismatches 24; Indels 2; Gaps 2;

DB 358 LIPIDEXYTTMRLAMEE-PCGPVLPITRTSVSEALISNK-SEYQASIFNDP 415
 QY 94 VFCOILAEBCVPAPPEPAPNAPASLAPTPSPVLEPNTSEPSDYALDSTFLOQHP 153

DB 416 RAF 418
 QY 154 AAF 156

RESULT 15
 ID A1AB_HUMAN STANDARD; PRT; 519 AA.
 AC P35368;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE ALPHA-1B ADRENERGIC RECEPTOR.
 GN ADRA1B.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 EC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93016158.
 RA RAMARAO C.S., DENKER J.M., PEREZ D.M., GATVIN R.J., RIEK R.P.,
 RA GRAHAM R.M.,
 RL J. BIOL. CHEM. 267:21936-21945(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 94239386.
 RA FORRAY C., BARD J.A., WETZEL J.M., CHIU G., SHAPIRO E., TANG R.,
 RA LEPOR H., HARTIG P.R., WEINSHANK R.L., BRANCHER T.A.,
 RA GLUCHOWSKI C.,
 RL MOL. PHARMACOL. 45:703-708(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95114877.
 RA SCHWAIN D.A., JOHNSTON G.I., PAGE S.O., MOSLEY M.J., WILSON K.H.,
 RA WORMAN N.P., CAMPBELL S., FIDOCK M.D., FURNESS L.M.,
 RA PARRY-SMITH D.J., PETER B., BAILEY D.S.,
 RL J. PHARMACOL. EXP. THER. 272:134-142(1995).
 CC -1- FUNCTION: THIS ALPHA-ADRENERGIC RECEPTOR MEDIATES ITS ACTION BY
 ASSOCIATION WITH G PROTEINS THAT ACTIVATE A PHOSPHATIDYLINOSITOL-
 CALCIUM SECOND MESSENGER SYSTEM.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 EMBL: M99589; -; NOT_ANNOTATED_CDS.
 EMBL: U03865; G494983; -;
 DR EMBL: L31773; G666891; -;
 DR PIR: A45121; A45121.
 DR GCRDB: GCR_0530; -;
 DR MIM: 104220; -;
 DR PROSITE: PS00237; G-PROTEIN RECEPTOR; 1.
 DR G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
 KW MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
 FT DOMAIN 1 45
 FT TRANSSEM 46 70
 FT DOMAIN 71 83
 FT TRANSSEM 84 105
 FT DOMAIN 106 115
 FT TRANSSEM 116 141
 FT DOMAIN 142 161
 FT TRANSSEM 162 182
 FT DOMAIN 183 201
 FT TRANSSEM 202 224
 FT DOMAIN 225 295
 FT TRANSSEM 296 319
 FT DOMAIN 320 326
 FT EXTRACELLULAR (POTENTIAL).

FT TRANSSEM 327 340 7 (POTENTIAL).
 FT DOMAIN 341 519 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 370 379 POLY-ARG.
 FT CARBOHYD 10 10-
 FT CARBOHYD 24 24 POTENTIAL.
 FT CARBOHYD 29 29 POTENTIAL.
 FT CARBOHYD 34 34 POTENTIAL.
 FT DISULFID 118 195 BY SIMILARITY.
 FT LIPID 365 365 PALMITATE (POTENTIAL).
 FT CONFLICT 379 379 MISSING (IN REF. 1).
 FT CONFLICT 497 500 AAAD -> PRH (IN REF. 1).
 FT CONFLICT 370 370 R -> RG (IN REF. 2).
 SQ SEQUENCE 519 AA; 56778 MW; D088058C CRC32;

Query Match 8.8%; Score 98; DB 1; Length 519;
 Best Local Similarity 41.3%; Pred. No. 6.05e-01;
 Matches 19; Conservative 9; Mismatches 15; Indels 3; Gaps 3;

DB 451 ALLSPAPPPGRRG-RHDSGP-LFTFKLLEPES-PGTDGASNG 493
 QY 12 TILQAPTAPSTIPGPRGSGPEITFTDPLPEPAPAPGPRASRG 57

Search completed: Fri Apr 17 13:38:42 1998
 Job time : 10 secs.


```

OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EC16;
RX MEDLINE: 93054355.
RA LINDBERG M., COLLIER A.;
J. BACTERIOL. 174:7385-7397(1992).
CC -1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLLOCATION OF THE
MULTIPLE PEPTIC ENZYMES.
CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE EXEC/PULC/OUTC FAMILY.
DR EMBL: L02214; G148433; -.
DR PIR: A47021; A47021.
DR PROSITE: PS01141; T2SP-C; 1.
KW TRANSPORT; TRANSMEMBRANE; INNER MEMBRANE.
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
TRANSMEM 17 35 POTENTIAL.
FT DOMAIN 36 272 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 272 AA; 30301 MW; 7944F46D CRC32;

Query Match
Best Local Similarity 36.7%; Pred. No. 4,56e-01;
Matches 22; Conservative 11; Mismatches 24; Indels 3; Gaps 3;

Db 6 LPLPSVIRILFYLLMLLFCQOL-AMIFWVGLP-DNSP-VASVOITPAQARQOPVTL 62
75 LPVEEPNPAKRLFLFLTLTYFCQILMAEGVPAPLPEDAPNMAASLAPTPVSPVLEPFNL 134

RESULT 12
ID GSOC_ERWCH STANDARD; PRT; 272 AA.
AC 001564;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE GENERAL SECRETION PATHWAY PROTEIN C (PEPTIC ENZYMES SECRETION PROTEIN
OUTC).
GN OUTC.
OS ERWINIA CHRYSANTHEMI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-3937;
RX MEDLINE: 93086427.
CONDEMNINE G., DOREL C., HUGOUVIEUX-COTTE-PATTAT N., ROBERT-BAUDOUY J.;
MOL. MICROBIOL. 6:3199-3211(1992).
CC -1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLLOCATION OF THE
MULTIPLE PEPTIC ENZYMES.
CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE EXEC/PULC/OUTC FAMILY.
DR EMBL: X65265; G42201; -.
DR PIR: S28013; S28013.
DR PROSITE: PS01141; T2SP-C; 1.
KW TRANSPORT; TRANSMEMBRANE; INNER MEMBRANE.
FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
TRANSMEM 17 35 POTENTIAL.
FT DOMAIN 36 272 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 272 AA; 30162 MW; 25634B57 CRC32;

Query Match
Best Local Similarity 36.7%; Pred. No. 6,05e-01;
Matches 22; Conservative 11; Mismatches 24; Indels 3; Gaps 3;

Db 6 LPLPSVIRILFYLLMLLFCQOL-AMIFWVGLP-DNSP-VASVOITPAQARQOPVTL 62
75 LPVEEPNPAKRLFLFLTLTYFCQILMAEGVPAPLPEDAPNMAASLAPTPVSPVLEPFNL 134

RESULT 13
ID VGLI_HSV11 STANDARD; PRT; 390 AA.

```

```

AC P06487;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE GLYCOPROTEIN I.
GN GI OR US7.
OS HERPES SIMPLEX VIRUS (TYPE 1 / STRAIN 17).
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85160822.
RA MCGEOCH D.J., DOLAN A., DONALD S., RIXON F.J.;
J. MOL. BIOL. 181:1-13(1985).
CC -1- THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND 2; GH, GB, GC,
GG, GD, GI, AND GE.
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEINS I, TO VZV GP1V,
AND TO PRV GP63.
DR EMBL: L00036; G291497; -.
DR EMBL: X14112; G58565; -.
DR EMBL: X02138; G59879; -.
DR PIR: A05243; Q0BE77.
KW GLYCOPROTEIN.
FT CARBOHYD 156 156 POTENTIAL.
FT CARBOHYD 175 175 POTENTIAL.
FT CARBOHYD 257 257 POTENTIAL.
SQ SEQUENCE 390 AA; 41369 MW; 7DA38E2D CRC32;

Query Match
Best Local Similarity 28.4%; Pred. No. 6,05e-01;
Matches 25; Conservative 21; Mismatches 37; Indels 5; Gaps 4;

Db 206 QASTPSTTSTSTIPASTTIPAPQASTP-FPTGDKPPQPGVNHPPSNA-TRATR 263
3 HSRCSPTTILQAPTPAST-IPGRSGCEIFTFDLPAPAPAGPSPASGHRKR 61

Db 264 DSR-YALTVTQIQIAPASIIAYEL 289
62 SRRVLPVRRVRLQPLVEEPNPAKRLFL 89

RESULT 14
ID GAPN_STMMU STANDARD; PRT; 475 AA.
AC 059931;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE NADP-DEPENDENT GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.9)
DE (NON-PHOSPHORYLATING GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE)
DE (GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (NADP+)) (TRIOSEPHOSPHATE
DEHYDROGENASE).
GN GAPN.
OS STREPTOCOCCUS MUTANS.
OC PROKARYOTA; FIRMICUTES; COCCI; STREPTOCOCCACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MG5 SEROTYPE C;
RX MEDLINE: 95270576.
BOYD D.A., CVITKOVIICH D.G., HAMILTON I.R.;
J. BACTERIOL. 177:2622-2627(1995).
CC -1- FUNCTION: IMPORTANT AS A MEANS OF GENERATING NADPH FOR
BIOSYNTHETIC REACTIONS.
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + NADP(+) + H(2)O
= 3-PHOSPHO-D-GLYCERATE + NADPH.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL: L18521; G642667; -.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; FALSE_NEG.
KW OXIDOREDUCTASE; NADP.
FT NP_BIND 230 235 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 250 250 BY SIMILARITY.
FT ACT_SITE 284 284 BY SIMILARITY.
SQ SEQUENCE 475 AA; 51221 MW; A621BB01 CRC32;

```


RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BOFIELD J., BUTTON J., CONNELL M., COPEX T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HARKINS T., HILLIER L., JIER M.,
RA JOHNSON L., JONES M., KERSHAW J., KISTEN J., LAISTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIKKEN L., ROOPER A., SAUNDERS B., SHOWNKEEN R.,
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULISTON J., THIERRY-MIEG J., THOMAS K., VAUDIAN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOLDMAN P.,
RL NATURE 368:32-38(1994).
DR EMBL: L16621; G289783; -.
DR PIR: S44920; S44920.
DR MORMPER: ZK688_5; CE00463.
DR PROSITE: PSS0053; UBIQUITIN²; UNKNOWN¹.
KW HYPOTHETICAL PROTEIN.
SEQUENCE 1799 AA: 202641 MW: 35F7DDB2 CRC32:

Very Match	9.88; Score 109; DB 1; Length 1799;
------------	-------------------------------------

Db	844	FLENRGRPRSTSSAPSRSENPPGFSFNS-EDAAIRGRGLPLGTRPPRRRTREVEVHAAA	902
		13 ILQAPTPASRIPGPRRSGCEIITFFDPLPPAAPAPGR-PSASRGHRKRSRYLYPRVY	71
Db	903	ARAESPNHISLTFTATHTTFAFAGF-PLUMASSNVPSRSGPPCWPILRQVVSPTPTRG	961
QY	72	RR-QQVPEEPNPARLFLFLTLTIVCQILMAEGVAPALP-PEDAPANAASLAPTPVSPVL	129
Db	962	FEFDLSSGSSDQ	972
QY	130	EPFNLTSPPSD	140

FT	DNA	751	755	POLY-SER.	
SEQ	SEQUENCE	1206	AA; 133464	MM; 6D70C261	CRC32; .
Query Match					
Best Local	Similarity	39.1%;	Score 103;	DB 1;	Length 1206;
Matches	18; Conservative	15;	Mismatches	9;	Indels 4; Gaps 4;

```

Db      716 VIALPNSSGGPPP-PPPPPPPPPGIAPP-PPGL-SFGSSSSSQY 758
      : : : : : : : : : : : : : : : : : : : : : : : :
Qy      98 IL-MAEGVPAPLPEDAPNAASLAPPVSVLPEPNTLSPSSDYA 142

```

CC -1- TISSUE SPECIFICITY: IT IS FOUND THROUGHOUT THE EMBRYO BUT
CC HAS A FUNCTIONAL ROLE ONLY IN THE KIDNEY AND LIMB.
CC -1- DEVELOPMENTAL STAGE: THIS IS THE ISOFORM FOUND IN THE APICAL
CC ECTODERMAL RIDGE AND THE MESENCHYMAL COMPARTMENT OF THE DEVELOPING
CC LIMB BUD.
CC -1- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE
CC PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A VARIATION IN
CC SPLICING IS SEEN AMONG DIFFERENT TISSUES AND DIFFERENT SIZE
CC TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE. THIS IS ISOFORM 4 AND IS
CC DIFFERENT IN ITS N-TERMINAL TO THE OTHER MOUSE ISOFORMS SO FAR
CC DETERMINED.
CC EMBL: X62379; G51553; -.
DR PIR: S24407; S24407.
DR HSSP: P19999; ICLG.
KW NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING.
FT DOMAIN 635 638 POLY-SER.
FT DOMAIN 644 744 PRO-RICH.

Db	61	RSRRVLYRVVRRQLPVEEPPAPARLFLLLTITVFCOILMAEEGVAPLPEDAPNASTL	120
Oy	61	RSRRVLYRVVRRQLPVEEPPAPARLFLLLTITVFCOILMAEEGVAPLPEDAPNASTL	120
Db	121	APTVPSPVLEPFNLTSPPSDYALDLSTFLQOHPAAF	156
Oy	121	APTVPSPVLEPFNLTSPPSDYALDLSTFLQOHPAAF	156
RESULT	2	STANDARD:	PRT: 153 AA.
ID	1EX1-MOUSE		
AC	P4694;		
DT	01-NOV-1995 (REL. 32, CREATED)		
DT	01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)		
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)		
DE	RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY		
DE	PROTEIN GLY96) (IMMEDIATE EARLY RESPONSE 3 PROTEIN).		
GN	IEX1 OR GLY96 OR IER3.		
OS	MUS MUSCULUS (MOUSE).		
OJ	EUKARYOTA, METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;		
	EUTHERIA; RODENTIA.		
	(1)		
AC	SEQUENCE FROM N.A.		
RC	STRAIN-BALB/C;		
RX	MEDLINE: 93117526.		
RA	CHARLES C.H., YOON J.K., SIMSKE J.S., LAU L.F.;		
RL	ONCOGENE 8:797-801(1993).		
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).		
CC	-1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE LUNG, TESTES		
CC	AND THE UTERUS.		
CC	-1- INDUCTION: BY SERUM GROWTH FACTORS.		
CC	-1- PPM: GLYCOSYLATED.		
CC	-1- SIMILARITY: STRONG, TO HUMAN ORTHOLOG.		
DR	EMBL: X67644; -: NOT_ANNOTATED_CDS.		
DR	PIR: S33363; S33363.		
DR	MCD: MGI:104814; IER3.		
KW	GLYCOPROTEIN; TRANSMEMBRANE.		
FT	DOMAIN 1 85 CYTOPLASMIC (POTENTIAL).		
FT	TRANSMEM 86 102 POTENTIAL.		
FT	DOMAIN 103 153 EXTRACELLULAR (POTENTIAL).		
FT	CARBOHYD 137 137 POTENTIAL.		
SO	SEQUENCE 153 AA; 16875 MW; 7A284EC2 CRC32;		
Query Match	67.5%; Score 752; DB 1; Length 153;		
Best Local Similarity	70.6%; Pred. No. 1,15e-120;		
Matches 108; Conservative	22; Mismatches 19; Indels 4; Gaps 2.		
Dh	1	MCHSRNHLHTGTGRLNAPSPASTGPELRGSGPEITFDPLPERAVASTARLNTSRGHRK	60
	1	MCHSRSCPTMTITLQAPTPAPSTIGPRRSGCEITFDPLPEPAAAPAGRPSASRGRK	60
Db	61	RSRRVLYRVVRRQLPVEEPPAPARLFLLLTITVFCOILMAEEGVAPLPEDAPNASTL	120
Oy	61	RSRRVLYRVVRRQLPVEEPPAPARLFLLLTITVFCOILMAEEGVAPLPEDAPNASTL	120
Db	121	EPISAPITAPVPLEPLNTSESSDYALDLAKFL	153
Oy	121	AP--TPVS--PVLEPFNLTSPPSDYALDLSTFL	149
RESULT	3	STANDARD:	PRT: 474 AA.
ID	VTB3-TTIVY		
AC	P19275;		
DT	01-NOV-1990 (REL. 16, CREATED)		
DT	01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)		
DT	01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)		
DE	VIRAL PROTEIN TPX.		
OS	THERMOPROTEUS TENAX VIRUS 1 (STRAIN VTB3) (TTV).		
CC	VIRIDAE; DS-DNA ENVELOPED VIRUSES; LIPOTHRIXVIRIDAE.		
GN	(1)		
RC	SEQUENCE FROM N.A.		
RX	MEDLINE: 90245666.		

[illegible]

W
E
S
E
R
E
H
(TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997, University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

rch_pp protein - protein database search, using Smith-Waterman algorithm
on: Fri Apr 17 13:38:32 1998: MasPar time 6.24 Seconds
626.917 Million cell updates/sec
Tabular output not generated.

Title: >US-08-799-910-10
Description: (1-156) from US08799910.pep
Perfect Score: 111.4
Sequence: 1 MCHSRSCHPMTITLQAPTPA.....EPSDYLDSTFLQHPAAF 156

Scoring table:
PAM 150
Gap 11

Searched: 69112 seqs, 25083644 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot35
1:swiss1

Statistics: Mean 44.412; Variance 91.621; scale 0.485

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Alt	Score	Query	Match	Length	ID	Description	Pred. No.
No.							
1	1102	98.9	156	1	IEX1_HUMAN	RADIATION-INDUCIBLE IM	5.97e-191
2	752	67.5	153	1	IEX1_MOUSE	RADIATION-INDUCIBLE IM	1.15e-120
3	117	10.5	474	1	VTR3_TTVIV	VIRAL PROTEIN TPX.	2.07e-03
4	112	10.1	228	1	VCOW_ADEMI	MINOR CORE PROTEIN (PR	9.76e-03
5	109	9.8	1799	1	YC25_CAEEL	HYPOPHYSICAL 202.6 KD	2.43e-02
6	103	9.2	1206	1	FOR4_MOUSE	FORMIN 4 (LIMB DEFORMI	1.44e-01
7	103	9.2	1468	1	FOR4_MOUSE	FORMIN 4 (LIMB DEFORMI	1.44e-01
8	101	9.1	268	1	CEBD_RAT	CCAAT/ENHANCER BINDING	2.57e-01
9	101	9.1	417	1	ALAB_CANFA	ALPHA-1B ADRENERGIC RE	2.57e-01
10	101	9.1	816	1	OAIF_NEUCR	OUTINIC ACID UTILIZATIO	2.57e-01
11	99	8.9	272	1	GSRC_ERWCH	GENERAL SECRETION PATH	4.56e-01
12	98	8.8	272	1	GSRC_ERWCH	GENERAL SECRETION PATH	4.56e-01
13	98	8.8	390	1	VGLI_HSV11	GLYCOPROTEIN 1.	6.05e-01
14	98	8.8	475	1	GAPN_STRMU	NADP-DEPENDENT GLYCERA	6.05e-01
15	98	8.8	519	1	ALAB_HUMAN	ALPHA-1B ADRENERGIC RE	6.05e-01
16	98	8.8	975	1	CDP_CANFA	CCAAT DISPLACEMENT PRO	6.05e-01
17	97	8.7	1043	1	CHS2_PABBR	CHITIN SYNTHASE 2 (EC	8.02e-01
18	97	8.7	1239	1	V120_EBV	CAPSID ASSEMBLY PROTEI	8.02e-01
19	97	8.7	1337	1	PTP4_HUMAN	PROTEIN-TYROSINE PHOSP	8.02e-01
20	96	8.6	245	1	VE4_HPV5B	PROBABLE E4 PROTEIN.	1.06e+00
21	96	8.6	245	1	VE4_HPV5B	PROBABLE E4 PROTEIN.	1.06e+00
22	96	8.6	1233	1	NME3_HUMAN	GLUTAMATE (NMDA) RECEP	1.06e+00
23	95	8.5	234	1	GLNA_DUNSA	GLUTAMINE SYNTHETASE (1.40e+00

RESULT	ID	IEX1_HUMAN	STANDARD;	PRT;	156 AA.
AC	P46695; Q93044;				
DT	01-NOV-1995 (REL. 32, CREATED)				
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)				
DE	RADIATION-INDUCIBLE IMMEDIATE-EARLY GENE IEX-1 (IMMEDIATE EARLY				
DE	PROTEIN GLY96).				
GN	IEX1.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;				
OC	EUTHERIA; PRIMATES.				
NC	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-PLACENTA;				
RX	MEDLINE: 96181295.				
RA	KONDRATYEV A.D., CHUNG K.N., JUNG M.O.;				
RL	CANCER RES. 56:1498-1502(1996).				
CC	[2]				
CC	PRELIMINARY SEQUENCE OF 1-106 FROM N.A.				
CC	TISSUE-PLACENTA;				
CC	HILLIER L., CLARK N., DUBOUE T., ELLISTON K., HAWKINS M., HOLMAN M.,				
CC	HUTTMAN M., KUCABA T., LE M., LENNON G., MARRA M., PARSONS J.,				
CC	RITKIN L., ROHLING T., TAN F., TREVASKIS E., WATERSTON R.,				
CC	WILLIAMSON A., WOHLMANN P., WILSON R.;				
CC	SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.				
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).				
CC	-1- INDUCTION: BY RADIATION.				
CC	-1- SIMILARITY: STRONG. TO MOUSE ORTHOLOG.				
CC	-1- CAUTION: REF. 2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO				
CC	FRAMESHIFTS.				
CC	EMBL: S81914; G1488385; -.				
CC	EMBL: T49531; -; NOT ANNOTATED_CDS.				
CC	GLYCOPROTEIN: TRANSMEMBRANE.				
CC	DOMAIN 1 82				
CC	DOMAIN 2 99				
CC	DOMAIN 3 100				
CC	DOMAIN 4 133				
CC	CARBOHYD 133				
CC	SEQUENCE 156 AA; 16973 MW; 3BD528CD CRC32;				

Query Match	Score	1102;	DB 1;	Length	156;
Best Local Similarity	98.7%				
Matches	154;	Conservative	1;	Mismatches	1;
Indels	0;	Gaps	0;		

1 MCHSRSCHPMTITLQAPTPASTIPGRPRGSGPEITFDPLPEPAAPAGRPSSGRHRK 60
|||||
1 MCHSRSCHPMTITLQAPTPASTIPGRPRGSGPEITFDPLPEPAAPAGRPSSGRHRK 60

Oy	1	ATGTGTACTCTCGACAGTCGCACACCACCATCACAATCCCTTGAGGGCCCCGAGCCC	60
Db	772	CCCTTCACACGGGCCCGGAATCTCGGGGGGCTGTGTGCCGAAATTTCACCTTGCACCT	831
Oy	61	CCCTTCACCAATCCCGGGGACCCCGGGGGCTCCGGCTCTGAGATCTTCACTTCACCTGACCT	120
Db	832	CTCCGGAGGGGGCGGTGTGTCCACCGGCCGTTTAAACTTCTTCGGGGGACCGAANA	891
Oy	121	CTCCGGAGGCCCGAGGGCCCTCTGCGGGGCCCCAGCGCTCTTCGGGGGACCGAANA	180
Db	892	CGCAGCCGAAGGTGTCTTACCTCGAGTGGT	923
Oy	181	CGCAGCCGCAAGGTTCTTACCTCGAGTGGT	212
RESULT	2		
LOCUS	RNPRG1	1758 bp	DNA
DEFINITION	R.norvegicus PRG1 gene.		
ACCESSION	X66437		
NID	91515318		
KEYWORDS	PRG1 gene.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; vertebrates; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 1758) Schäfer H., Trauzold A., Siegel E.G., Folsch U.R. and Schmidt W.E. PRG1: a novel early-response gene transcriptionally induced by pituitary adenylate cyclase activating polypeptide in a pancreatic carcinoma cell line Cancer Res. 56 (11), 2641-2648 (1996)		
JOURNAL	Cancer Res. 56 (11), 2641-2648 (1996)		
MEDLINE	96221139		
REFERENCE	2 (bases 1 to 1758)		
AUTHORS	Trauzold A.		
TITLE	Direct Submission Submitted (05-MAR-1996) A. Trauzold, Laboratory of Molecular Gastroenterology, 1st Dept. of Medicine, University of Kiel, Schlittenheimsstrasse 12, Kiel, D-24105, FRG location/Qualifiers 1. 1758 /organism="Rattus norvegicus" /db_xref="taxon:10116" /cell_line="AR4-2J" 590..1279 /gene="PRG1" 590..1279 /gene="PRG1"		
FEATURES			
source	1. 1758		
BASE COUNT	358 a 358 c 470 g 455 t		
ORIGIN			
Query Match	9.3% Score 114; DB 14; Length 1758;		
Best Local Similarity	76.9%; Pred. No. 3,336-46;		
Matches 153; Conservative	0; Mismatches 49; Indels 0; Gaps 0;		
Db	590 ATGTGCACTTCGGATAACCACTTCACACCATGACTGAGGCTCGGGCCGACGTT	649	
Oy	1 ATGTGCACTTCGGACACTCCACCCGACCATGACATCTCTCAGGCCGCCGCC	60	
Db	650 CCGCTCACCGGCCCGAATCTCGGGGGGCTCCGGTCCCGGAATTTTACCTTCGACCC	709	
Oy	61 CCCCTCACCAATCCCGGGGACCCCGGGGGCTCGGTCTGAGATCTTCACCTTCGACCT	120	
Db	710 CTCCGGAGGGGGGAGGTGTGTCCACCGGCCGTTTAAACTTCTTCGGGGGACCGAANA	769	
Oy	121 CTCCGGAGGCCCGAGGGCCCTCTGCGGGGCCCCAGCGCTCTTCGGGGGACCGAANA	180	
Db	770 CGCAGCCGAGGTTCTTACCTCGAGTGGT	801	
Oy	181 CGCAGCCGCAAGGTTCTTACCTCGAGTGGT	212	
RESULT	3		

LOCUS	166494	7218 bp	DNA	PAT	23-DEC-1997
DEFINITION	Sequence 14 from patent US 5670367.				
ACCESSION	166494				
NID	g2724471				
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified. 1 (bases 1 to 7218) Donner,F., Scheiflinger,F. and Falkner,F.Gunter. TITLE Recombinant fowlpox virus JOURNAL Patent: US 5670367-A 14 23-SEP-1997; FEATURES Location/Qualifiers 1..7218				
BASE COUNT	1944 a	1491 c	1486 g	1929 t	368 others
ORIGIN					
Query Match	6.4%;	Score 79;	DB 25;	Length 7218;	
Best Local Similarity	1.3%;	Pred. No. 4,27e-26;			
Matches	5;	Conservative 220;	Mismatches 146;	Indels 0;	Gaps 0;
Dn	1061 TGGCATTYYY	1120			
Oy	99 TGAGATTCTTCACTTCGACCCTCCCGAGCCCGACGCCCTCCCGGCAGCCCAG	158			
Dn	1121 YY	1180			
Oy	159 GCGCTCTGCGGGCACCAAGCCGACGCGAGGTCTTACCTCGATGCTGCGCGC	218			
Dn	1181 YY	1240			
Oy	219 CCAAGTCGCACTCGAAGAACCAACCAAGCCAAAGGCTCTTTCTGCTGCTACCAT	278			
Dn	1241 YY	1300			
Oy	279 CGCTCTGCGCAGATCTGATGCTGAAGAGGTGTCGCGCGCTGCTCCAGAGA	338			
Dn	1301 YY	1360			
Oy	339 GCGCCCTAAGCCGATCCCTGCGGCCACCCCTGTGCCCCGCTCTCGAGCCCTTAA	398			
Dn	1361 YY	1420			
Oy	399 TCTGACTCGAGGCCCTCGAGCTACGCTCTGAGACTCAGACACTTCTCCAGACAACC	458			
Dn	1421 YYYYYYYYYY 1431				
Oy	459 GGCGCGCTCT 469				
RESULT	4	7218 bp	DNA	PAT	23-DEC-1997
LOCUS	166494				
DEFINITION	Sequence 14 from patent US 5670367.				
ACCESSION	166494				
NID	g2724471				
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified. 1 (bases 1 to 7218) Donner,F., Scheiflinger,F. and Falkner,F.Gunter. TITLE Recombinant fowlpox virus JOURNAL Patent: US 5670367-A 14 23-SEP-1997; FEATURES Location/Qualifiers 1..7218				
BASE COUNT	1944 a	1491 c	1486 g	1929 t	368 others
ORIGIN					
Query Match	5.5%;	Score 67;	DB 25;	Length 7218;	
Best Local Similarity	2.4%;	Pred. No. 1.82e-19;			
Matches	9;	Conservative 214;	Mismatches 156;	Indels 0;	Gaps 0;

CDS		join(2328..2426,6476..6690,6751..7462)
	/gene="Cp1"	
	/codon_start=1	
	/product="cysteine proteinase-1"	
	/db_xref="pid:g2305221"	
	/translation="MRTAAIPLLALLAANAQSFADVYMEEHNTFKLEHRKNODEEERRAKIFENKKRIKAKHNORPREGVSFKLAVNKYADLLHERROLNGNFTYLHQURADESEFKGVTEISPAHVTLPRSVDMRTKGAVTVXKQCGSCGMFSSGALLEGHFERSKGYLVLSLEQNLDSCSTKYGNNGCCNAAGVDNAFRYIKNGJIDIEKSPYEAIIDSCHEKGTVGATDRGPTDIPQDEERKKMAEVAIVGVPSVALIDASHEPOFSEGVYEPPQCDAOLDHGLVLVGFGEDESGEDYWLTKNSMTGWGDCKFIKLNRKEQGCIASASSPYLV"	
intron	2427..6475	
	/gene="Cp1"	
misc_feature	4546..4553	
	/gene="Cp1"	
	/note="insertion site of P(Caspen)(50C)"	
	/citation=[1]	
exon	6476..6690	
	/gene="Cp1"	
	/number=3	
intron	6691..6750	
	/gene="Cp1"	
exon	6751..7207	
	/gene="Cp1"	
	/number=4	
mRNA	join(<8110..9300,9370..>9532)	
	/product="phenylalanyl tRNA synthetase"	
CDS	join(8110..9300,9370..>9532)	
	/note="potential orf"	
	/codon_start=1	
	/product="phenylalanyl tRNA synthetase"	
	/db_xref="pid:g2305222"	
	/translation="MLTTRRVGGARHMLSTRCLASSAPAKSPSSPOLESGSTVDTGMYNPKILSYVGANKRLDTHDPLSLTIROIVAYFGVARNGRNPFSYDDMGAEVYQQNQEPDLIRPADYVRQKSDCYIINQOHLRAHTTAQVLYIGGLDNFLPEYRRDELDSHTHPVFHQADVRLVTRDKLFERNPELEFEETWSGTIADLPFLILPHASESWTPNSPVATRAAVKLMHEMKHVLVGLTKDFGRIRYRWVDYFPPTOPMELETFKDWMSELCCGIMRHREIIQRSGVHSIGYAFGVGLERLAVLDPIDIRLFNSNDSEGLSOFSEKDLHNLPRYKPISHYPQCTNDLSFWLPQDIEDVAGFSPNFYDVRVSAGVAGFDELISLVDFKFKPKTKGSVCFRIVYRMERTLQAENVNIHKQIASASVDSEFNQIX"	
BASE COUNT	2929 a 2357 c 2282 g 3046 t	158 others
ORIGIN		
Query Match	2.9%: Score 35; DB 17; Length 10772;	
Best local similarity 19.0%; Pred. No. 3.32e-03;		
Matches 20; Conservative 55; Mismatches 28; Indels 2; Gaps 2;		
Dn 1700 TTTMMKMWTMTKMAAMKTYRTMMKKMYMSRTTTTSAMMMYTWTSTWTK [*] YMAAYAMN 1758		
: : : : : : : : : : : : :		
Cp 809 TTGCATAAGTTAAATAATATAATTACGTACA-CATCTCATCACCTAGAGACGTACATA 751		
Dn 1759 KMMWTTTMAAMASHARKWKTSAAAAYSAMRKMMKWMPRAWK 1803		
:: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::		
Cp 750 AATACATATAAATATTAAATTAGACATAGAATAAATAATTACG 706		
RESULT 8	128278	215 bp DNA PAT 30-OCT-1996
LOCUS	Sequence 5 from patent US 5569830.	
DEFINITION	128278	
ACCESSION	g1819054	
NID	91819054	
KEYWORDS	Unknown.	
SOURCE	Unknown.	
ORGANISM	Unclassified.	
REFERENCE	1 (bases 1 to 215)	
AUTHORS	Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.	
TITLE	Plant inhibitors of fungal polygalacturonases and their use to control fungal disease	
JOURNAL	Patent: US 5569830-A 5 29-Oct-1996;	
FEATURES	Location/Qualifiers	

	source	1..215	/organism="unknown"	source						
BASE COUNT	15	a	8	c	25	g	26	t	141	others
ORIGIN										
Query Match		2.7%	Score 33;	DB 25;	Length 215;					
Best Local Similarity		12.2%;	Pred. NO. 2,61e-02;							
Matches	25;	Conservative	79;	Mismatches 101;	Indels 0;	Gaps 0;				
Dd	11	VVSRTASCNDKAKKDNGNTSSMTDCCNRTGVCVDJDTTHVYNNDSGNHKKSSANYVG	70							
Oy	799	AACCTATACCAGGGGTGGAGATGTCCTCGTGTGAATGCACGCTCTTGATTAT	858							
Dd	71	NNVCAAKHYHTNVSGADSKTYTDSYNASGTSSNGITDGNRSGADSYGSKTYMTSR	130							
Oy	859	TGAGCTTTGTGGGACTGTGTGACMCAGGACAACCTGGAACCTCGGCCAAGAATGAGAA	918							
Dd	131	NRTGKTANNAYDSRRMGDAASYGSDKNTKKHAKN\$ADGVSKKNNGDNRNRRTGTSTKS	190							
Oy	919	ATGGGAGAGACTCGGCTGGGGAGAGACCTCCGGCTGGAGTAAGTAGTCTGCGTGGG	978							
Dd	191	NNCGGANKRDVSSYANNKC CGSSCT	215							
Oy	979	TAACTTTAGGAGGTGACTGCATCCT	1003							

LOCUS	9	800 bp	mRNA	VRT	02-APR-1996
DEFINITION	GGU41467				
ACCESSION	Gallus gallus fibroblast growth factor 8 FGFB mRNA, partial cds.				
KEYWORDS	g1134969				
SOURCE	chicken.				
ORGANISM	Eukaryotes; mitochondrion eukaryotes; Metazoa; Chordata; Vertebrates; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.				
REFERENCE	1 (bases 1 to 800)				
AUTHORS	Crossley, P.H., Minowada, G., MacArthur, C.A. and Martin, G.R.				
TITLE	Roles for FGFB in the induction, initiation, and maintenance of chick limb development				
JOURNAL	Cell 84 (1), 127-136 (1996)				
MEDLINE	96140646				
REFERENCE	2 (bases 1 to 800)				
AUTHORS	Crossley, P.H., Minowada, G., MacArthur, C. and Martin, G.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (27-NOV-1995) Philip H. Crossley, Anatomy, UCSF, Parnassus, San Francisco, CA 94143-045, USA				
FEATURES	Location/Qualifiers				
source	1. 800				
	/organism="Gallus gallus"				
	/db_xref="taxon:9031"				
	/dev_stage="embryo"				
	/tissue_type="brain"				
	/clone_lib="E10 embryonic chick brain cdNA library"				
	<1. 615				
	/codon_start=1				
	/product="fibroblast growth factor 8 FGFB"				
	/db_xref="FID:g1134970"				
	/translation="VEMHLFVLCQAQVTVSPNFTQHVREOSLVTDLSRLVRYT				
	QVSTSGKHVQILDNKKINAMAEQGDVHAFLIVETDFGSRVRKGAATFYICMNR				
	KGKIGKNGKGDCEVPEIYLENNYTVLQNAKEGWMATRGPRKSGSKTHORH				
	EVHFKKRLPKOHOTREPHRFEELEPPRRNSKRTRNSARVP"				
BASE COUNT	225 a 219 c 226 g 130 t				
ORIGIN					
Query Match	2.4%;	Score 29;	DB 16;	Length 800;	
Best Local Similarity	74.6%;	Pred. NO. 1.36e+00;			
Matches	44;	Conservative 0;	Mismatches 15;	Indels 0;	Gaps 0;
Db	742 TGTTCGCTTTTACAAAAAACCACCAACAAACAAAAAAGCTGAG 800				
	1170 TGTTCGTGAACCAATAATGCTTTTACCTGCTCAAAAAAAGCTGAG 1228				

LOCUS	10	RA	MAV	11-JAN-1995
DEFINITION	SUSL36LBP 1125 bp			
ACCESSION	S.scrofa mRNA for L-36 lactose binding lectin.			
NID	X79303			
KEYWORDS	g623345			
SOURCE	1-36lbp gene; lactose-binding lectin.			
ORGANISM	pig.			
REFERENCE	Sus. scrofa			
AUTHORS	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Artiodactyla; Suidae; Sus.			
TITLE	Chiu M.L., Parry D.A., Feldman S.R., Klapper D.G. and O'Keefe E.J. (bases 1 to 1125)			
	An adherens junction protein is a member of the family of lactose-binding lectins			
JOURNAL	J. Biol. Chem. 269 (50), 31770-31776 (1994)			
MEDLINE	95081129			
REFERENCE	2 (bases 1 to 1125)			
AUTHORS	O'Keefe, E.J.			
TITLE	Direct Submission			
JOURNAL	Submitted (18-MAY-1994) E.J. O'Keefe, University of North Carolina			
FEATURES	137 NCMH, Dept of Dermatology, CB7600, Chapel Hill, NC 27514, USA			
SOURCE	Location/Qualifiers			
	1..1125			

	gene	/db_xref="taxon:9823"	
		/dev_stage="adult"	
		/tissue_type="epithelium, oral, tongue"	
		/cell_type="keratinocyte"	
		/clone_id="pig tongue cDNA"	
	49..1020		
	/gene="1361bp"		
	49..1020		
	/gene="1361bp"		
	/codon_start=1		
	/product="L-36 lactose binding protein"		
	/db_xref="PID:g623346"		
	/translation="MVFAPAGIOTPTNPILPYKPIPGGIRVGMYSYITGYANHHM RFENFVVGQPGADVAFHEPNRFDGWKGVFNSSQQDSKMGNEEKRSMPRKAAFT LVINWLEDEHKVVVNGDPFEYEFGRIPQVLTHLOVDDELISINFIGQPAPSPGG MPNKGPGPKHNOQPNLPCEMGATPVPVKTRLOGGLVARIVTIVIGYPPSG KSLINFRKVGSSGVVALHINRLREGIYVRNSYLNGKAERKSEFMFPAGQFYDD SIRCLDFRYKYANGOHFLPDFSHLSNQGVDTEIQGDVLTLSYQI"		
BASE COUNT	255 a 347 c 303 g 220 t		
ORIGIN			
Query Match	2.4%; Score 29; DB 15; Length 1125;		
Best Local Similarity	89.2%; Pred. No. 1.36e+00;		
Matches	33; Conservative 0; Mismatches 4; Indels 0; Gaps 0;		
Dd 1089	AATAAATTATTAACCTGCCAAAAAAAATTTAAAAA 1125		
-Oy 1185	AATAAATTGATTACTGTCAAAAAAAAATTTAAAAA 1221		
LOCUS	11 BOVIOPBP 1266 bp mRNA MAM 25-JAN-1993		
DEFINITION	Bovine inorganic pyrophosphatase mRNA sequence.		
ACCESSION	M95283		
NTD	g163228		
KEYWORDS	inorganic pyrophosphatase.		
SOURCE	Bos taurus retina cDNA to mRNA.		
ORGANISM	Bos taurus		
	Eukaryotes; mitochondrion eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos. 1 (bases 1 to 1266) Yang,Z., and Wensel,T.G. Molecular cloning and functional expression of cDNA encoding a mammalian inorganic pyrophosphatase J. Biol. Chem. 267, 24641-24647 (1992)		
REFERENCE	JOURNAL MEDLINE 93077559		

[illegible]

```

REFERENCE
AUTHORS      Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Rodentia; Sciuromorphia; Muridae; Murinae; Mus.
TITLE         1 (bases 1 to 1479)
              Penney, M., Wilkinson, C., Wallace, M., Javeratz, J.P., Seeger, M.,
              Dubel, W., Toda, T., McKay, S., Allshire, R. and Gordon, C.
              Multidrug resistance is mediated by the 26S proteasome in fission
              yeast
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 1479)
AUTHORS      Gordon, C.
TITLE        Direct Submission
              Submitted (08-MAY-1997) C. Gordon, MRC Human Genetics Unit,
              Developmental Genetics, Western General Hospital, Crewe Road,
              Edinburgh EH4 2XU, UK
FEATURES
SOURCE       Location/Qualifiers
              1..1479
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /dev_stage="adult"
                237..1166
                /gene="PAD1"
                237..1166
                /gene="PAD1"
                /function="proteolysis"
                /codon_start=1
                /product="26S proteasome, non-ATPase subunit"
                /db_xref="PID:e352083"
                /db_xref="PID:g2505940"
                /translat="MDRLRLRGGMPLARPPTDAPAVDAEOYVYSSILLLKMLKKH
              RAGPMVEVMGLMGEFDDYDTVRYIDYFAPOSGTGVSEAVDPFQAKMDIKQNG
              RPEVMYGVSHRPGFGLSGVDINTQGSFALSERVAVVYVDIYQYKGVYDARFR
              LINANMVLGHERPQITSNLGLNKPSTQLIHGLNRHYISITINRKNLEQMLLN
              LHKRSMEGLTLDYSECHKHNSVSEMLELAKNYKAAVEEDKMPQDLAIKNGVK"
              QDPRRLHEEHVYDLMETNIVQICLAAMDYVFK"
BASE COUNT   453 a          270 c          349 g          407 t
ORIGIN
Query Match   2.4%; Score 29; DB 14; Length 1479;
Best Local Similarity 87.2%; Pred. No. 1.36e+00;
Matches 34; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db 1440 AATTAATACATTACATCCAAAAA1478
Cy 1184 AATTAATGATTACTGCAAAAAA1222
RESULT 14
LOCUS      XU069669      1811 bp      mRNA      VRT      26-FEB-1997
DEFINITION Xenopus laevis nuclear pore complex-associated protein TPR (tptr)
            mRNA, partial cds.
ACCESSION   U69669
NID         g1850343
KEYWORDS    African clawed frog.
SOURCE      African clawed frog.
ORGANISM    Xenopus laevis
            Eukaryote; metazoan; chordate; vertebrate;
            Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
            Pipidae; Xenopodidae; Xenopus.
REFERENCE   1 (bases 1 to 1811)
AUTHORS     Cordes, V.C., Reidenbach, S., Ratzlitz, H.R. and Franke, W.W.
TITLE        Identification of protein p270/7pr as a constitutive component of
            the nuclear pore complex-attached intranuclear filaments
            J. Cell Biol. 136 (3), 515-529 (1997)
            97177132
            2 (bases 1 to 1811)
            Cordes, V.C., Hunziker, A. and Franke, W.W.
            Direct Submission
            Submitted (06-SEP-1996) Cell Biology/0110, German Cancer Research
            Center, INF 280, Heidelberg 69120, Germany
            Location/Qualifiers
              1..1811
                /organism="Xenopus laevis"
                /db_xref="taxon:8355"

```

gene /tissue-type="kidney"
/clone_id="Uni-ZAPMXR library, Stratagene"
1.1678
/gene="tpc"
<1.1678
/gene="tpc"
/note="nuclear pore complex-associated protein;
translocated promotor region"
/codon_start=2
/product="tpc"
/db_xref="PID:91850344"
/translation="VPTQSHATIESPTQETPVEIYVSSVERPPTTSITGTSARP
SSSIKPRPREEDSTLETPEQIADDDQRTKKEEDIEKTEAVINTEALH
LTCQSMNEFPLEEIVSPTQSVIESQAPQOLNQSTQSDLPKTHMLVIV
ISDENEDQGEDEEDDEEDAGIEGDSNETSADGNEDEGDAEAD
GTDDPTEDSTAGEGNORADSONIGDGVTAESFTQETREOPSSASDRGPRP
POSPRORAHPRRLITLAPROLOGPAPRAIRIVAROSGRGLQITPGVGRQHFEDE
DRIVSPFTLVPHRIDGFAEAIHSPOVAGVPRFRGPEEDMPQASSHSDGLASQ
GGLMDYDPLFLAHEESGSRVPTTLOVAPVSVFAENPADTSDHASOSVPMVTT
STGNVPTSVDSGADEGEVEVEASEGIGASTLEMDTQCEEPVOPSEADLPSTSD
PSSSIADTSSSKPRRWMLQPOGGRPFKRSRGSDFRGRGINSNI"

BASE COUNT 577 a 427 c 427 g 380 t

ORIGIN

Query Match 2.4%; Score 29; DB 16; Length 1811;
Best Local Similarity 78.4%; Pred. No. 1.36e+00;
Matches 40; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 1761 TTCTATGACGAAATTAATCTTTATTTAAAAA 1811
Oy 1172 TTCTGTGAACAAATAATGATTACTGTCAAAAAA 1222

RESULT 15

LOCUS DDU73686 1840 bp DNA INV 08-DEC-1997

DEFINITION Dictyostelium discoidium cytosolic glycoprotein FP21 (fpa2) gene,
complete cds.

ACCESSION U73686

NID 91658023

KEYWORDS Dictyostelium discoidium.

SOURCE Dictyostelium discoidium

ORGANISM Eukaryote; Dictyostelida; Dictyostelium.

REFERENCE 1 (bases 1 to 1840)

AUTHORS West,C.M., Kozarov,E. and Teng-umnuay,P.

TITLE The cytosolic glycoprotein Fp21 of Dictyostelium discoidium is
encoded by two genes resulting in a polymorphism at a single amino
acid position

JOURNAL Gene 200 (1-2), 1-10 (1997)

MEDLINE 98038971

REFERENCE 2 (bases 1 to 1840)

AUTHORS West,M.C., Kozarov,E. and Teng-umnuay,P.

TITLE Direct Submission

JOURNAL Submitted (08-OCT-1996) Anatomy & Cell Biology, University of
Florida, 1600 SW Archer Rd., Gainesville, FL 32610-0235, USA

FEATURES

location/Qualifiers

1.1840

/organism="Dictyostelium discoidium"

/strain="Ax3"

/db_xref="taxon:44689"

1061.1703

/gene="fpa2"

join(1061..1157,1312..1703)

/gene="fpa2"

/note="SKP1-like; similar to the product of the fpa1 gene"

/product="cytosolic glycoprotein Fp21"

/db_xref="pid:91658024"

/translation="NSLVKLESSDEKVEPIEKELACMSVTIKNMIEDIGESDAPILP
NVTSTILKLVDCRHHOHPSGGDDKDEKLDIPYDRDFCKVDPTLFELILA
ANYLDIKPLDLVCTKTVANMIRKTPPEIRKIFINIKNDFTPEEQRKEMWECDDG
GN"

1158..1311

Intron

BASE COUNT 742 a 166 c 170 g 762 t

ORIGIN

Query Match 2.4%; Score 29; DB 17; Length 1840;
Best Local Similarity 87.2%; Pred. No. 1.36e+00;
Matches 34; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 915 TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT 953
Cp 1222 TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT 1184

Search completed: Thu Apr 23 05:25:29 1998
Job time : 1010 secs.

 W O R L D
 (TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1997 University of Edinburgh, U.K.
 Distribution Rights by Oxford Molecular Ltd

MPsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Apr 23 05:39:41 1998: Maspar time 157.52 Seconds

Tabular output not generated.

Title: >US-08-799-910-9

Description: (1-1228) from US08799910.seq

Perfect Score: 1228

N.A. Sequence:

Comp: 1 ATGTCACACTCTCCGACGCTG.....AAAAAAAAAACTCGAG 1228
 TACACAGTGAAGAGCGTCGAC.....TTTTTTTTTTTGAGCTC

Scoring table: TABLE default

Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 159651 seqs, 57698962 bases x 2

Post-processing: Minimum Match 0#

Listing first 45 summaries

Database:

n-geneseg30
 1:part1 2:part2 3:part3 4:part4 5:parts 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33

Statistics: Mean 9.360; Variance 6.778; scale 1.381

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	135	11.0	297	10	Human genome fragment	1.11e-57
2	94	7.7	1047	2	Human Natriuretic Pep	5.93e-35
3	80	6.5	1047	2	Human Natriuretic Pep	2.20e-27
4	62	5.0	267	10	Human genome fragment	6.80e-18
5	47	3.8	33	18	Human gene signature	2.61e-10
6	46	3.7	91	9	Oligonucleotide probe	8.07e-10
7	43	3.5	204	1	Base substituted E.co	2.31e-08
8	43	3.5	204	1	Base substituted E.co	2.10e-07
9	41	3.3	114	12	Generic DNA sequence	1.59e-05
10	37	3.0	114	12	Generic DNA sequence	1.32e-04
11	35	2.9	114	12	Generic DNA sequence	4.59e-05
12	36	2.9	114	12	Generic DNA sequence	4.59e-05
13	36	2.9	114	12	Generic DNA sequence	4.59e-05
14	36	2.9	114	12	Generic DNA sequence	4.59e-05
15	35	2.9	178	31	Human endothelin-1 an	1.32e-04

C	16	34	2.8	100	31	T76186	Human IL4 receptor an	3.74e-04
C	17	34	2.8	114	12	070465	Generic DNA sequence	3.74e-04
C	18	34	2.8	114	12	070466	Generic DNA sequence	3.74e-04
C	19	34	2.8	114	12	070470	Generic DNA sequence	3.74e-04
C	20	34	2.8	114	12	070465	Generic DNA sequence	3.74e-04
C	21	34	2.8	128	31	T76233	Human IL6 antisense o	1.05e-03
C	22	33	2.7	114	12	070469	Generic DNA sequence	1.05e-03
C	23	33	2.7	178	31	T76405	Human endothelin-1 an	2.93e-03
C	24	32	2.6	114	12	070467	Generic DNA sequence	2.93e-03
C	25	32	2.6	114	12	070468	Generic DNA sequence	2.93e-03
C	26	32	2.6	172	32	T76363	Human Interleukin 8 a	8.08e-03
C	27	31	2.5	114	12	070470	Generic DNA sequence	8.08e-03
C	28	31	2.5	2420	7	Q46672	p mysa cDNA clone.	5.92e-02
C	29	29	2.4	88	31	T76170	Human IL3 receptor an	2.20e-02
C	30	30	2.4	100	31	T76186	Human IL4 receptor an	2.20e-02
C	31	30	2.4	114	12	070473	Generic DNA sequence	2.20e-02
C	32	30	2.4	114	12	070471	Generic DNA sequence	2.20e-02
C	33	30	2.4	114	12	070472	Generic DNA sequence	2.20e-02
C	34	29	2.4	190	31	T76452	Chymase antisense o1	5.92e-02
C	35	29	2.4	250	31	T76438	Substance P antisense	1.57e-01
C	36	28	2.3	114	12	070471	Generic DNA sequence	1.57e-01
C	37	28	2.3	130	31	T76152	Human vascular cell a	1.57e-01
C	38	28	2.3	162	31	T76307	Human RANTRIS antisens	1.57e-01
C	39	28	2.3	200	31	T76398	Human leukotriene C4	1.57e-01
C	40	28	2.3	595	19	T16988	E-Dex integrin inhibi	1.57e-01
C	41	28	2.3	2179	13	Q73117	Alpha 2, 3-sialyl tra	1.57e-01
C	42	28	2.3	2232	12	Q77739	Human alpha-2,3-sialy	1.57e-01
C	43	28	2.3	2992	13	Q79934	Human liver hap cDNA.	1.57e-01
C	44	28	2.3	2992	20	T06491	Human hepatoma cellno	1.57e-01
C	45	28	2.3	2992	11	Q65572	Human liver hap cDNA.	1.57e-01

ALIGNMENTS

RESULT	ID	Query Match	Score	DB ID	Length	Mismatches	Indels	Gaps
1	077534 standard; DNA; 297 BP.	11.0%	135	DB 10	297	2	2	2
AC	077534:	Best Local Similarity	76.6%	Pred. No. 1.11e-57;				
DT	23-SEP-1994 (first entry)	Matches	209	Conservative	0	Mismatches	62	Indels
DE	Human genome fragment. (Preferred)							
KW	Brain; Placenta; bone marrow; genetic analysis; gene mapping;							
KM	detection; homology; human; adrenal tissue; ds.							
OS	Homo sapiens.							
PN	WO9401548-A.							
PF	20-JAN-1994.							
PR	13-JUL-1993; G01467							
PA	13-JUL-1992; GB-014657.							
PI	(MED1-) MEDICAL RES COUNCIL.							
PI	Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;							
PI	Sibson DR, Starkey M;							
DR	MP1: 94-035056/04.							
PT	New nucleic acid fragment encoding gene products - can be used							
PS	for genetic analysis and mapping							
PS	Claim 1; Page 575-576; 61pp; English.							
CC	Human nucleic acid fragments, isolated from brain, adrenal tissue,							
CC	the placenta or bone marrow comprise any of: (A) a sequence							
CC	selected from (076401-077613), (B) an allelic variation of a							
CC	sequence as described in (A), or (C) a sequence complementary							
CC	to (A) or (B).							
CC	Preferred sequences exhibit no more than 90105mology to a human							
CC	sequence known per se.							
SQ	Sequence 297 BP: 66 A; 117 C; 61 G; 52 T;							
DB	Query Match	11.0%	Score 135	DB 10	Length 297;			
DB	Best Local Similarity	76.6%	Pred. No. 1.11e-57;					
OY	7 CACTTCGCGAGCCGACCCGACGACATCTCAGGCCGCCGCCGCCCTCC	66						
DB	82 gttatcgcagccacaataag-clygcgcctcagatctccacccctcagacccctccg	140						
OY	67 ACCATCCCGGAGCCCGGGGGGCGCTCGGATGATCTACCTTCGACCCCTCCCG	126						

Db 141 aagccacagcgccctccgagcaccagcttc-ctcgagcagcaaaagcgagc 199
 127 GACCCGCGAGCGCCCTCCGCGCCGCGCCGCGCTCTCGCGGCGCAAGACCGCAGC 186
 Oy 200 tccaggaatctcaacctagatcgagcgagcttcaggttaaggagcaaaccc 259
 187 CGCAGGGTCTCTACCTCGAGTGTGCGCGCCGCGCAGCTGCGAGTGAAGAACCGCA 246
 Db 260 gccagaagctcttcctcggtgatcgatc 292
 247 GCCAAAGGCTCTCTTCTGCGCTCACCACATC 279
 Oy

RESULT 2
 ID Q10572 standard; DNA; 1047 BP.

AC Q10572;

DE 09-APR-1991 (first entry)

DE Human Natriuretic Peptide Receptor B.

KM NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;

KM hyperaldosteronism; glaucoma; guanylyl cyclase.

OS Homo sapiens.

Key Location/Qualifiers

FT Peptide 1..22

FT /label- signal sequence 12

FT /label- mature NPBR 23..455

FT /label- extracellular domain

FT /note- "binds natriuretic peptides A, B and C"

FT Domain 456..456

FT /label- transmembrane domain

FT Domain 479..1047

FT /label- cytoplasmic domain

FT /note- "GC and protein kinase activity"

FT Modified-site 24..26

FT /label- N-glycos-site 35..37

FT /label- N-glycos-site 161..163

FT /label- N-glycos-site 195..197

FT /label- N-glycos-site 244..246

FT /label- N-glycos-site 277..279

FT /label- N-glycos-site 349..351

FT /label- N-glycos-site 600..602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

FT /label- N-glycos-site 602

Query Match 7.7%; Score 94; DB 2; Length 1047;
 Best Local Similarity 9.0%; Pred. No. 5,93e-35;
 Matches 83; Conservative 263; Mismatches 568; Indels 10; Gaps 10;

Db 16 gvimngarntnavvnnhnnhnsyawwvrvnvanavanangrnnvndrnvsnnn 75
 Oy 304 GAAAGAGGTGTGCGCGCCCTCCCTCCCAAGAGACCCCTTAACGCGCATCTGCGG 363
 Db 76 snynanasaavdnknyhdnnngvcvynasvsnrnsnshwznnnrtagavnsakh 135
 Oy 364 CCGACCCCTGTGTCCCGCTCCGAGCCCTTAATGTGACTGGAGCCCTCGGA-CTA 422
 Db 136 yrtvrtgsanckngnvnrtngjnnwkarannndartddrnhnrtngvnnangsn 195
 Oy 423 CGCTGTGACCTCAGACCTTCTCCAGCAACACCGCGCCCTTCACTGATGATCC 482
 Db 196 nsvnhvvarngnngnnaethnrtangrvnvcgnnnnnnnnnnnnnnnnnngdyvnn 255
 Oy 483 CGCACTCCCAAAAAGAAATCCGAAAACACACAAAAGAAACACACAGCGCTGCTGCGG 542
 Db 256 dvngsnragnrtatgrnndrtlnnananrnnnnrtvnnrtvnnnnnnnnnnnnnn 315
 Oy 543 AGAGCGTATCCCACTGAGCTCCGAGCA-ACTGAACTGAAACACTACAGCGAG 601
 Db 316 rarnngvngnngsnnnnngcnydnnnyannnnnnnnnnngtrndgnrvnkmgrryhg 375
 Oy 602 AGCCACCCGGTGTGAGCGGAGCGAGCGACAGAGACCGAGCGGATACAGACCG 661
 Db 376 vtgnvmdkndrntdvnnvnmgdndgdnnaahysganknw-tgrnnvkygans 434
 Oy 662 AGGCACAGCCAGCTGAGCGGCTGAGCGGCTGAGAGAGAGCGCTGTAATTTATTTCT 721
 Db 435 dnnncandndnscklnstnanvngtgnltnmpvssnnnrknnnnknaasm-wr 493
 Oy 722 ATGTCTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 781
 Db 494 nrvnnnnngsnryhkgagsrn-tnsrsgssysnmlahkyannanltgnkgnvankh 552
 Oy 782 TACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840
 Db 553 vnktrnnnrtvnnnnkhrdvnnnhltnngacndnnncvntnycnrsndnnnds 612
 Oy 841 AGTCTCTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 899
 Db 613 nndnmnysnndnvgmannhnsnnsnscvdsrvnhtdygnasnst 672
 Oy 900 CGCAGAACTAGCAAGAAATG-GGAGAGACTCGGAGGAGAGACTCCCGCTGGGA 958
 Db 673 nndnnaayakknnlannnnsgnnnttgmnaadvysngnnnnnnnnnnnnnnnnng 732
 Oy 959 TGAAGT-CTGTTGTG-GGTGTGTAAGTTTGAAGAGTACAGCACTCCAGCACTCAAC 1016
 Db 733 nknvnkvngnrvnynnsndrtlnnnnnvnnnnrcwandnandnngnknrvnkn 792
 Oy 1017 TCCGCTGTCTACTGTGTGAGACTTCGCGGAGCAATAGAAATGATCCGTGAGATCT 1076
 Db 793 gntsnndnn 852
 Oy 1077 TCATCTCTTGAAGTGGCTTTAGGCTGCTGCAAGTGAAGGCTTGGGCTTGGG 1136
 Db 853 nanandsvtnysdnvntansanstnnvntnnndnntcndandndnykvtngday 912
 Oy 1137 CTGTCAGAGCAGCACTGCGAGATCGGCTGATAGTCTGTGGAACAACAATTAATGAT 1196
 Db 913 mvsgnngnrgnrtannanmana 936
 Oy 1197 TTAAGTCAAAAAAATAAAAAA 1220

RESULT 3
 ID Q10572 standard; DNA; 1047 BP.
 AC Q10572;
 DE 09-APR-1991 (first entry)

```

DE Human Natriuretic Peptide Receptor B.
KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW hyperaldosteronism; glaucoma; guanyl cyclase.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= signal sequence
FT Protein 12
FT /label= mature NPRB
FT Domain 23..455
FT /label= extracellular domain
FT /note= "binds natriuretic peptides A, B and C)"
FT Domain 456..456
FT /label= transmembrane domain
FT Domain 479..1047
FT /label= cytoplasmic domain
FT /note= "GC and protein kinase activity"
FT Modified-site 24..26
FT /label= N-glycos-site
FT Modified-site 35..37
FT /label= N-glycos-site
FT Modified-site 161..163
FT /label= N-glycos-site
FT Modified-site 195..197
FT /label= N-glycos-site
FT Modified-site 244..246
FT /label= N-glycos-site
FT Modified-site 277..279
FT /label= N-glycos-site
FT Modified-site 349..351
FT /label= N-glycos-site
FT Modified-site 600..602
FT /label= N-glycos-site
PN MO9100292-A.
PD 10-JAN-1991.
PF 22-JUN-1990; 003586.
PR 23-JUN-1989; US-370673.
PA (GERTH ) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
DR WPI: 91-036711/05.
DR N-PSDB: Q10324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PS kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3: Fig 1: 49pp: English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prepd.
SQ Sequence 1047 BP: 87 A; 15 C; 83 G; 51 T;

Query Match 6.5%; Score 80; DB 2; Length 1047;
Best Local Similarity 8.5%; Pred. No. 2.20e-27;
Matches 75; Conservative 248; Mismatches 546; Indels 10; Gaps 10;

```

```

CP 700 CTCCTTCCACCGGCGCTAGCCCAAGCTGGCTGCTCTCTATGCCCTCGCTC 641
DB 268 ratg-rnmndtrnmananranntvnrtnymnnnnnnnnnnnnnnnnnnnnnnnn 326
CP 640 TCTGTGCGGCTGGTCCCGCTCAGACACCGGGTGGCGCTCCGCTGTGTCTGAGT 581
DB 327 nsmnnnagcnydygnnnnyannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 385
CP 580 TCAGATTGCTCGGAAGTCCAGTAGTGGATAGCTCTCGCCACACAGTACGCTGGTG 521
DB 386 ndrntnrvnwangdndsgdnnaahysgnknmwrgyrnnrvkgsnnndnnncandnd 445
CP 520 TTTCTTGTGGTTTTCGGATTCTTTTGGGAGTGGGGGAGTCACTTAGAAGCGGG 461
DB 446 nscdktnsnanvanganvgnntnmogvssnnnnrkmnnknnasmmwrtrvnnnnnqsn 505
CP 460 CCGGGTGTCTCGAGAGAAAGTCTGAGTCCAGACCGTAGTCCAGGGCTCCGAAGTCA 401
DB 506 ryhka-gsrntnsrg-saysnmtahkxymnnantghknvvanhknvknrrnntn 563
CP 400 GATTAAAGGCTCTGAGGAGCGGGGACACAGGGGTGGCGCCAGGATCGCGGTTAGGG 341
DB 564 vnnnkhmrdrvnnhntnngacndnnnnenvnyenrnsndnnnnndnnnnnryen 623
CP 340 CGTCTCTGAGGACAGGCGCGCGCACACCTTTCAGCCATCAGAGTCTGGCAGAGA 281
DB 624 nndvngmnnhnsnshgnsncvdrnrvnknrtgynasnrstannndnnanayak 683
CP 280 CGATGTGACAGACA-GAAGAGAACCTTTTGGCTGGGTTC-TCGACTGGCAG 223
DB 684 knntannnsgnnnttgmnaadvysngnnnnnnnnnnnnnnnnnnnnnnnnnnnn 743
CP 222 CTGGCCCGGACACTCGAGGGTA-GAAGAACCTTGGCTGGCTTCGTGGCCCGGAG 164
DB 744 nrrynrnsndrtnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 803
CP 163 AGCGCTGGGGCGCCCGGAGGGCGCTCGCGGCTCCGAGAGGGTGAAGTGAAGA 104
DB 804 rnmnyannnnknvnnrtngannnkrkanannnnnnnnnnnnnnnnnnnnnnnnnn 863
CP 103 TCTCAGAGCCGAGACCCCGCGGGTCCGGGATGTGTGAGGGGCGCGGGTGGCGCT 44
DB 864 ncdvntansanstnnvntnnndhntcndannndv 902
CP 43 GCAGATGTCATGTCGGGTGGCTGACGTGGAAGTGC 5

RESULT 4
ID Q77554 standard; DNA: 267 BP.
AC Q77554;
DT 23-SEP-1994 (first entry)
DE Human genome fragment. (Preferred)
KW Brain; Placenta; Bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.
OS Homo sapiens.
PN MO9401548-A.
PD 20-JAN-1994.
PF 13-JUL-1993; G01467.
PR 13-JUL-1992; GB-014857.
PA (MED1-) MEDICAL RES COUNCIL.
PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
PI Sibson DR, Starkey M;
DR WPI: 94-035056/04.
PT New nucleic acid fragment encoding gene products - can be used
PT for genetic analysis and mapping
PS Claim 1: Page 586; 616pp: English.
CC Human nucleic acid fragments, isolated from brain, adrenal tissue,
CC the placenta or bone marrow comprise any of: (A) a sequence
CC selected from (Q76401-Q76613), (B) an allelic variation of a
CC sequence as described in (A), or (C) a sequence complementary
CC to (A) or (B).
CC Preferred sequences exhibit no more than 90105mology to a human
CC sequence known per se.

```

SQ Sequence 267 BP; 58 A; 117 C; 49 G; 42 T;
 Query Match 5.0%; Score 62; DB 10; Length 267;
 Best Local Similarity 75.0%; Pred. No. 6,80e-18;
 Matches 114; Conservative 0; Mismatches 35; Indels 3; Gaps 3;
 DB 116 cctgaagatctccacctaagaccctcccccgaagccgacccgcccctgc-gaaccttc 174
 |||||
 OY 97 CCGAGATCTTCACCTTCGACCTCTCCGAGACCCGACGCGCCCTCGCGGCGCC 156
 |||||
 DB 175 aactcgt-tcacagcgcgcaagacctatccaggactctcgcctcctagctga-cggn 232
 |||||
 OY 157 AGCGCCTCTCGCGGACCGCAAGCGCAGCGAGGTTCTCTACCTCGAGTGGTCCG 216
 |||||
 DB 233 cctcactcgtcgtcaggaatcggaactcgc 264
 |||||
 OY 217 CGCCAGCTCGCAGCTCGAGAACCGAACCCAGC 248
 |||||
 RESULT 5
 ID T20789 standard; cDNA to mRNA; 53 BP.
 AC T20789;
 DT 09-JUL-1996 (first entry)
 DE Human gene signature HUMGS02013.
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KM human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 OS Homo sapiens.
 PN W09514772-A1.
 PD 01-JUN-1995.
 PF 11-NOV-1994; J01916.
 PR 12-NOV-1993; JP-355504.
 PA (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.
 PA Matsubara K, Okubo K;
 PI WPI: 95-206931/27.
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 PS Claim 1; Page 723; 2245pp; Japanese.
 CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in T19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types
 SQ Sequence 53 BP; 18 A; 9 C; 9 G; 17 T;
 Query Match 3.8%; Score 47; DB 18; Length 53;
 Best Local Similarity 100.0%; Pred. No. 2.61e-10;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 gatcgctagtattgttcgtgaacaacaataattgatttcagtc 47
 |||||
 OY 1158 GATCGCTAGTATGTTCTGTGAACACAATAAATTGATTACTGTC 1204
 |||||
 RESULT 6
 ID O51746 standard; cDNA; 91 BP.
 AC O51746;
 DT 31-MAY-1994 (first entry)
 DE Oligonucleotide probe MK14-A
 KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
 KW

KW ss.
 OS Synthetic.
 PN EP-571911-A.
 PD 01-DEC-1993.
 PF 24-MAY-1993; 108325.
 PR 26-MAY-1992; US-889651.
 PA (BECT) BECTON DICKINSON CO.
 PI Shank DD, Spears PA;
 PI WPI: 93-378844/48.
 PT New oligo:nucleotide probes specific for Mycobacteria - used for
 PT detection and amplification of Mycobacteria nucleic acid in
 PT samples
 PS Claim 3; Page 14; 23pp; English.
 CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
 CC (O51735). It hybridized to all spp. of mycobacteria tested, but
 CC cross reacted to a few non-mycobacterial spp. The probe may
 CC be useful as an initial screen for mycobacterial infection.
 CC See also O51735-45 and O51747-59.
 SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;
 Query Match 3.7%; Score 46; DB 9; Length 91;
 Best Local Similarity 13.3%; Pred. No. 8.07e-10;
 Matches 8; Conservative 45; Mismatches 7; Indels 0; Gaps 0;
 DB 1 ggcctcgcgssvhsyvvhvshvshvshvshvshvshvshvshvshvshvshvsv 60
 |||||
 OY 88 GGCTCGGTCGAGATCTTCACTTCGACCTCTCCGAGCCCGAGCGGCGCCCTGTC 147
 |||||
 RESULT 7
 ID O51746 standard; cDNA; 91 BP.
 AC O51746;
 DT 31-MAY-1994 (first entry)
 DE Oligonucleotide probe MK14-A
 KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
 KW ss.
 OS Synthetic.
 PN EP-571911-A.
 PD 01-DEC-1993.
 PF 24-MAY-1993; 108325.
 PR 26-MAY-1992; US-889651.
 PA (BECT) BECTON DICKINSON CO.
 PI Shank DD, Spears PA;
 PI WPI: 93-378844/48.
 PT New oligo:nucleotide probes specific for Mycobacteria - used for
 PT detection and amplification of Mycobacteria nucleic acid in
 PT samples
 PS Claim 3; Page 14; 23pp; English.
 CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
 CC (O51735). It hybridized to all spp. of mycobacteria tested, but
 CC cross reacted to a few non-mycobacterial spp. The probe may
 CC be useful as an initial screen for mycobacterial infection.
 CC See also O51735-45 and O51747-59.
 SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;
 Query Match 3.5%; Score 43; DB 9; Length 91;
 Best Local Similarity 2.0%; Pred. No. 2.31e-08;
 Matches 1; Conservative 45; Mismatches 3; Indels 0; Gaps 0;
 DB 13 vhsyvvhvshvshvshvshvshvshvshvshvshvshvshvshvshvshvshvsv 61
 |||||
 CP 1153 CAGTCGCTCGCTGACAGCCACCAACCCACCCACTTACTCGACGC 1105
 |||||
 RESULT 8
 ID N81164 standard; DNA; 204 BP.
 AC N81164;
 DT 08-NOV-1990 (first entry)
 DE Base substituted E.coli beta-galactosidase alpha-fragment.
 KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
 OS Escherichia coli.
 FH key Location/Qualifiers
 FT misc_feature 19..69


```
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcripiase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions/ most of which
CC occurred singularly in any given mutant.
```

QY	Query Match	3.3%	Score 41:	DB 1:	Length 204:
Dd	Best Local Similarity	18.7%;	Pred. No. 2.10e-07;	Mismatches	29; Conservative 69; Mismatches 54; Indels 3; Gaps 2;
Dc	Sequence	204 BP;	21 A;	47 C;	17 G;
Df					11 T; 108 Others;
Dg	cyttcaatccgcgaagcgacbcyrragggtccccgggywcgagacygaaycdchnc	83	:	:	:
Dh	: : : : : :	:	:	:	:
Qz	CTCtCGACGACTGCCACCAGCACCATGCATCTGTCAAGCCGGCCCCGTC--	65	:	:	:
Dj	84 cgmmttbhyrhmhbnyrydnrsdaawycyrsvdydcymachdhvvybbby	143	:	:	:
Dk	66 CACCACTCCCAGGCCCGCGGGGGTCTCGGTCGTGAATCTTGACCTTGACCTTCTCC	125	:	:	:
Dl	144 nvhnanccebnhvchnvbhnmhrmwayrd	178	:	:	:
Dm	126 GGAGCCCCAGCGGCCCTCGCGGGCGGCCAAGCG	160	:	:	:
Result	10				
ID	Q70469 standard; DNA; 114 BP.				
DT	07-APR-1995 (first entry)				
DE	Generic DNA sequence to generate a random TSAR peptide library.				
KW	TSAR: totally synthetic affinity reagent; synthetic; binding domain;				
KS	effector domain; concatenated heterofunctional protein; linker;				
KV	direct; rapid; detection; screening; treatment; generic; ss.				
OS	Synthetic.				
FH	Key	Location/Qualifiers			
FT	misc-feature	55..60			
FT	/tag= a				
FT	sequence of 6,9 or 12 nucleotides (see				
FT	comments)"				
FM	WO9418318-A.				
PD	18-AUG-1994.				
PE	01-FEB-1994; UO00977.				
PR	01-FEB-1993; US-013416.				
PR	30-DEC-1993; US-176500.				
PR	31-JAN-1994; US-189331.				
PA	(UINC-) UNIV NORTH CAROLINA.				
PI	Fowles DM; Kay BK;				
PT	WP1: 94-279739/34.				
PT	Identifying proteins or peptide(s) which bind a ligand - by				
PT	screening a recombinant vector library expressing fusion proteins				
PS	comprising a binding domain and an effector domain				
PS	disclosure; Page 35; 25pp; English.				
CC	This formula is a generic DNA sequence used to generate random TSAR peptide				
CC	libraries. This formula can be represented as follows: X(TGC)(NNB)10-				
CC	(TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC).X and Y are flanking restriction				
CC	sites ('X' is not the same as 'Y') that are not specified further. This				
CC	sequence generates peptides that are cleavable in structure. Other				
CC	generic sequences are shown in Q70465-68. Other specific peptides				
CC	generated by these genetic sequences are shown in R65150-54. TSARS are				
CC	concatenated heterofunctional proteins or peptides, comprising at least				
CC	two functional regions - a binding domain with affinity for a ligand and				
CC	a second effector peptide portion that is chemically or biologically				
CC	active.They may further comprise a linker peptide between the 2 domains.				
CC	The oligonucleotides are also designed so that the expressed peptide				
CC	contains 2 or 4 cysteine residues positioned in, or flanking, the				
CC	unpredicted or variant residues.These residues confer some degree of				
CC	conformational rigidity to the peptides. The TSARS or compns.comprsing				
CC	a TSAR binding domain can be used in vivo to deliver a chemically or				
CC	biologically active moiety, eg. metal ion, radiotracer, peptide, toxin				

Query Match	2.9%;	Score 35;	DB 12;	Length 114;
Best Local Similarity	10.1%;	Pred. No. 1.32e-04;		
Matches	11;	Conservative	29;	Mismatches 69;
			Indels	0;
			Gaps	0

[illegible]

```

RESULT 13
ID 070467 standard: DNA: 114 BP.
AC 070467;
DE 05-APR-1995 (first entry)
KW Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc-feature 55..60
FT /tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
FT WO9418318-A.
PN 18-AUG-1994.
PD 01-FEB-1994: U00977.
PR 01-FEB-1993: US-013416.
PR 30-DEC-1993: US-176500.
PR 31-JAN-1994: US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR MPI: 94-279739/34.
DR P-PSDB: R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC 070467 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)15(TGC)(NNB)1Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in 070466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compens. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed actively allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.9%; Score 36; DB 12; Length 114;
Best Local Similarity 3.7%; Pred. No. 4.59e-05;
Matches 4; Conservative 33; Mismatches 71; Indels 0; Gaps 0;

Db 3 bnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnnnnnn 62
Cp 696 TTCCACGCGGCCCTAGCCCGGCTGCGTCTCTATGCGCCGCTGCTG 637
Db 63 bnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnb 110
Cp 636 TGCCTCTGCTGCGCTCCGCTCAAGACCGGGTGGCTCTCCGCTAGTG 589

```

```

RESULT 15
ID 070466 standard: DNA: 114 BP.
AC 070466;
DE 05-APR-1995 (first entry)
KW Generic DNA sequence to generate a random TSAR-9 peptide library.
KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc-feature 55..60
FT /tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
FT WO9418318-A.
PN 18-AUG-1994.
PD 01-FEB-1994: U00977.
PR 01-FEB-1993: US-013416.
PR 30-DEC-1993: US-176500.
PR 31-JAN-1994: US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR MPI: 94-279739/34.
DR P-PSDB: R65152.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC 070466 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)1(TGC)(NNB)10(TGC)2(NNB)8(TGC)(NNB)
CC -9Y. X and Y are flanking restriction sites (X is not the same as Y)
CC that are not specified further. Other generic sequences are shown in
CC 070466-68. Other specific peptides generated by these generic sequences
CC are shown in R65151-54. TSARs are concatenated heterofunctional proteins
CC or peptides, comprising at least two functional regions - a binding
CC domain with affinity for a ligand and a second effector peptide portion
CC that is chemically or biologically active. They may further comprise a
CC linker peptide between the 2 domains. The oligonucleotides are also
CC designed so that the expressed peptide contains 2 or 4 cysteine residues
CC positioned in, or flanking, the unpredicted or variant residues. These
CC residues confer some degree of conformational rigidity to the peptides.
CC The TSARs or compens. comprising a TSAR binding domain can be used in
CC vivo to deliver a chemically or biologically active moiety, eg. metal
CC ion, radioisotope, peptide, toxin or enzyme, to the specific target or
CC on the cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed actively allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;

Query Match 2.9%; Score 36; DB 12; Length 114;
Best Local Similarity 7.3%; Pred. No. 4.59e-05;
Matches 8; Conservative 31; Mismatches 71; Indels 0; Gaps 0;

Db 5 gcnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnnnnnnnn 64
Cp 703 GCTCTCTGCTGCGCTCCGCTCAAGACCGGGTGGCTCTCTATGCGCCCTCG 644
Db 65 nbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnnnnnnnnnn 114
Cp 643 GCTCTCTGCTGCGCTCCGCTCAAGACCGGGTGGCTCTCTCGCTG 594

```

M P S R C H P
(TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Apr 17 13:40:06 1998; Maspar time 10.54 Seconds
Tabular output not generated. 633.960 Million cell updates/sec

Title: >US-08-799-910-10
Description: (1-156) from US08799910.pep
Perfect Score: 1114
Sequence: 1 MCHSRSCHPMTITLQAPTPA.....EPSDVALDSTFLQOHPAAF 156

Scoring table: PAM 150
Gap 11

Searched: 195121 seqs, 42852602 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: plr55
1:plr1 2:plr2 3:plr3 4:plr4

Statistics: Mean 36.831; Variance 125.421; scale 0.294

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	1111	99.7	156	2	JC5537 This is a DE line.	8.18e-119
2	752	67.5	153	2	S33363 This is a DE line.	1.10e-74
3	117	10.5	474	2	S15921 This is a DE line.	1.63e-01
4	114	10.2	351	2	B34768 This is a DE line.	3.09e-01
5	110	9.9	228	2	S53504 This is a DE line.	7.13e-01
6	109	9.8	1729	2	MSTQOEAS This is a DE line.	8.78e-01
7	102	9.2	753	2	JU0532 This is a DE line.	3.67e+00
8	103	9.2	913	2	S20590 This is a DE line.	3.00e+00
9	102	9.2	1106	2	JQ0405 This is a DE line.	3.67e+00
10	103	9.2	1206	2	S24407 This is a DE line.	3.00e+00
11	103	9.2	1468	2	S11515 This is a DE line.	4.48e+00
12	101	9.1	268	2	B39429 This is a DE line.	4.48e+00
13	101	9.1	417	2	E30341 This is a DE line.	4.48e+00
14	101	9.1	816	2	F31277 This is a DE line.	4.49e+00
15	100	9.0	333	2	S61849 This is a DE line.	5.49e+00
16	100	9.0	439	2	S51939 This is a DE line.	5.49e+00
17	100	9.0	1492	2	A40333 This is a DE line.	6.70e+00
18	99	8.9	272	2	A47021 This is a DE line.	8.17e+00
19	98	8.8	272	2	S28013 This is a DE line.	8.17e+00
20	98	8.8	351	2	S50754 This is a DE line.	8.17e+00
21	98	8.8	390	1	QOBE77 This is a DE line.	8.17e+00
22	98	8.8	475	2	A57151 This is a DE line.	8.17e+00
23	98	8.8	516	2	JC2332 This is a DE line.	8.17e+00

24	98	8.8	517	2	A45121	This is a DE line.	8.17e+00
25	98	8.8	975	2	S33121	This is a DE line.	8.17e+00
26	98	8.8	1486	2	B40333	This is a DE line.	8.17e+00
27	97	8.7	1239	1	QOBE10	This is a DE line.	9.95e+00
28	97	8.7	1337	1	I38670	This is a DE line.	9.95e+00
29	97	8.7	1337	2	I52599	This is a DE line.	9.95e+00
30	96	8.6	245	1	W4WLS	This is a DE line.	1.21e+01
31	96	8.6	245	1	W4WLB5	This is a DE line.	1.21e+01
32	96	8.6	356	2	E70032	This is a DE line.	1.21e+01
33	95	8.5	234	1	AJDHO	This is a DE line.	1.47e+01
34	95	8.5	316	2	G02424	This is a DE line.	1.47e+01
35	95	8.5	336	2	S55611	This is a DE line.	1.47e+01
36	95	8.5	377	2	A48018	This is a DE line.	1.47e+01
37	95	8.5	449	2	S16748	This is a DE line.	1.47e+01
38	95	8.5	493	2	JH0158	This is a DE line.	1.47e+01
39	95	8.5	534	2	S21961	This is a DE line.	1.47e+01
40	95	8.5	643	2	S55610	This is a DE line.	1.47e+01
41	95	8.5	3866	2	B48205	This is a DE line.	1.47e+01
42	95	8.5	3869	2	A48205	This is a DE line.	1.47e+01
43	94	8.4	232	2	JU0069	This is a DE line.	1.79e+01
44	94	8.4	389	2	S27200	This is a DE line.	1.79e+01
45	94	8.4	473	2	S50755	This is a DE line.	1.79e+01

ALIGNMENTS

RESULT	1	STANDARD;	PRT:	156 AA.
ID	JC5537			
XX	xxxxxx			
AC	xxxxxx			
DT	01-JAN-1900			
XX				
DE	This is a DE line.			
XX				
CC	A:Experimental source: monocyste			
CC	A:Note: the authors translated the codon CCG for residue 106 as Arg			
SO	SEQUENCE 156 AA; 16914 MW; 127043 CN;			

Query Match 99.7%; Score 1111; DB 2; Length 156;
Best Local Similarity 99.4%; Pred. No. 8.18e-119;
Matches 155; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db	1	MCHSRSCHPMTITLQAPTPASTIPGRSGPEITFDPLPEPAAPAGPSGRGRK 60
Db	61	RSRRVLPVVRRLPYEENPAPAKRLFLTLTIVFCOILMAEGVPAAPLPEDAPNASTL 120
Qy	1	MCHSRSCHPMTITLQAPTPASTIPGRSGPEITFDPLPEPAAPAGPSGRGRK 60
Qy	61	RSRRVLPVVRRLPYEENPAPAKRLFLTLTIVFCOILMAEGVPAAPLPEDAPNASTL 120
Db	121	APTPVSPVLEPNTSEPSDVALDSTFLQOHPAAF 156
Qy	121	APTPVSPVLEPNTSEPSDVALDSTFLQOHPAAF 156
RESULT	2	
ID	S33363	STANDARD; PRT: 153 AA.
XX	xxxxxx	
AC	xxxxxx	
DT	01-JAN-1900	
XX		
DE	This is a DE line.	
XX		
CC	A:Cross-references: EMBL:X67644	
CC	C:Genetics: 70/3	
CC	A:Introns: 70/3	
CC	C:Keywords: transmembrane protein	
SO	SEQUENCE 153 AA; 16875 MW; 122478 CN;	

Query Match 67.5%; Score 752; DB 2; Length 153;
Best Local Similarity 70.6%; Pred. No. 1.10e-74;

OY 135 TSEPS 139

RESULT 6
ID MSETQEOASGNGEPDLPPTIRVLTGLDREAVTIGLQDTIQSLIDGRREMIQSGFORYIAGRVL STANDARD

AC xxxxxx
XX 01-JAN-1900
DT This is a DE line.
DE
XX
SO SEQUENCE 1729 AA; 194986 MW; 14801419 CN;

Query Match 9.8%; Score 109; DB 2; Length 1729;
Best Local Similarity 23.7%; Pred. No. 8.78e+01;
Matches 31; Conservative 30; Mismatches 65; Indels 5; Gaps 5;

Db 774 FLENGRIPSTSSAPSTSENPGPFENS-EDADIRAGRLPLGTRPNRRTVRETVHPAAA 832
OY 13 ILQAPTPAPSTIPGRRSGPEITFTDPLPEPAAPAGR-PSASGHHKRSRVLYPRVY 71
Db 833 ARAESPNIHSLTFTATHTFAPAGF-PLMASSNVPSTSGPPGPIRQVVSPTPTBGL 891
OY 72 RR-QLPVEEPNPAKRLFLTLTIVFCQILMAEGVPAPLP-PEADPNASLASLAPTVSPVL 129
Db 892 FEFDLGSSDQ 902
OY 130 EFPNLTSEPSD 140

RESULT 7
ID JQ0532 STANDARD; PRT; 753 AA.

AC xxxxxx
XX 01-JAN-1900
DT This is a DE line.
DE
XX
XX A:Accession: JQ0532
CC A:Molecule type: mRNA
CC A:Residues: 1-753 <DIN>
CC A:Experimental source: strain Jervis Bay isolate
SO SEQUENCE 753 AA; 82427 MW; 3187938 CN;

Query Match 9.2%; Score 102; DB 2; Length 753;
Best Local Similarity 32.9%; Pred. No. 3.67e+00;
Matches 24; Conservative 14; Mismatches 30; Indels 5; Gaps 5;

Db 271 LPRPGNRGVLDPGPKYRGA-PSS-NL-PLPTPOAPPRAERLQRSILHSHASRONAPRLRP 327
OY 14 LQAPTPAPSTIPGPR-RGSGPEITFTDPLPEPAAPAGRPSASGHHKRSRVLYPRVY 72
Db 328 RRLR-SDPIQOTR 339
OY 73 ROLPVEEPNPAK 85

RESULT 8
ID S20590 STANDARD; PRT; 913 AA.

AC xxxxxx
XX 01-JAN-1900
DT This is a DE line.
DE
XX
XX TOIG of: s20590 check: 4303 from: 1 to: 913
CC >P1:S20590
CC exo-alpha-stalidase (EC 3.2.1.18) - Actinomyces viscosus

CC C:Species: Actinomyces viscosus
CC C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 09-Sep-1997
CC C:Accession: S20590
CC R:Henningsen, M.; Roggentin, P.; Schauer, R.
CC Biol. Chem. Hoppe-Seyler 372, 1065-1072, 1991
CC C:A:Title: Cloning, sequencing and expression of the stalidase gene from Actin

CC C:A:Reference number: S20590; M01D:92162190
CC A:Accession: S20590
CC A>Status: preliminary
CC A:Molecule type: DNA
CC A:Residues: 1-913 <HEI>
CC A:Cross-references: EMBL:X62276; NID:q39254; PID:q39255
CC C:Keywords: glycosidase; hydrolase
SO SEQUENCE 913 AA; 96216 MW; 3970532 CN;

Query Match 9.2%; Score 103; DB 2; Length 913;
Best Local Similarity 29.9%; Pred. No. 3.00e+00;
Matches 20; Conservative 14; Mismatches 31; Indels 2; Gaps 2;

Db 816 SPASRNAAPTCKGMEPDEID-RPSDGTMAOPTGAP-ARRVPRRRRRRPAAGCLARDQ 873
OY 16 APTAPSTIPGRRSGPEITFTDPLPEPAAPAGRPSASGHHKRSRVLYPRVYRQL 75
Db 874 RAADPGP 880
OY 76 PVEEPNP 82

RESULT 9
ID JQ0405 STANDARD; PRT; 1106 AA.

AC xxxxxx
XX 01-JAN-1900
DT This is a DE line.
DE
XX
XX A:Accession: JQ0405
CC A:Molecule type: DNA
CC A:Residues: 1-1106 <SHI>
CC A:Cross-references: EMBL:X15867
CC A>Note: all the codons between two in-frame stop codons are translated; the
CC A>Note: the gene encoding this protein overlaps uvra gene
SO SEQUENCE 1106 AA; 119484 MW; 4705861 CN;

Query Match 9.2%; Score 102; DB 2; Length 1106;
Best Local Similarity 29.9%; Pred. No. 3.67e+00;
Matches 20; Conservative 16; Mismatches 29; Indels 2; Gaps 2;

Db 569 PGPAGVPAGRRPRVOPAPAGRHPLRRRGA-HPGHTDRLAGRRPLRRAVHRPAPA 627
OY 19 PAPSTIPGRRSGPEITFTDPLPEPAAPAGRPSASGHHKRSRVLYPR-VVRQLPV 77
Db 628 GQPEPRH 634
OY 78 EEPNPAK 84

RESULT 10
ID S24407 STANDARD; PRT; 1206 AA.

AC xxxxxx
XX 01-JAN-1900
DT This is a DE line.
DE
XX
XX TOIG of: s24407 check: 501 from: 1 to: 1206
CC >P1:S24407
CC form in isoform IV - mouse
CC C:Species: Mus musculus (house mouse)
CC C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 10-Sep-1997

 M P E R E H (TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1997 University of Edinburgh, U.K.
 Distribution Rights by Oxford Molecular Ltd

Mparch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 17 13:39:01 1998; Maspar time 12.21 Seconds

Tabular output not generated.

Title: >US-08-799-910-10

Description: (1-156) from US08799910.pep

Sequence: 1 MCHSRCHPTMTILQAPTPA.....EPDVALDLSTFLQHPAAAF 156

Scoring table: PAM 150

Gap 11

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 08

Listing first 45 summaries

Database:

sptrembl5
 1:sp_fungi 2:sp_human 3:sp_invertebrate 4:sp_mammal
 5:sp_mhc 6:sp_organelle 7:sp_phage 8:sp_plant
 9:sp_bacteria 10:sp_fodent 11:sp_virus 12:sp_vertebrate
 13:sp_unclassified

Statistics: Mean 42.617; Variance 98.163; scale 0.434

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1114	100.0	156	2	092691	8.46e-174
2	1114	100.0	351	11	085302	2.07e-02
3	1112	100.1	115	9	P95306	3.65e-02
4	1112	100.1	285	1	P78977	3.65e-02
5	1110	100.1	3247	11	065553	4.84e-02
6	1110	99.9	225	11	084630	6.41e-02
7	110	99.9	8	043558	PROLINE RICH PROTEIN P	6.41e-02
8	108	99.7	552	8	096343	1.12e-01
9	102	99.2	418	3	001662	5.74e-01
10	102	99.2	568	3	02712	5.74e-01
11	102	99.2	640	11	084171	5.74e-01
12	102	99.2	680	11	098187	5.74e-01
13	103	99.2	913	9	059164	4.39e-01
14	103	99.2	1711	9	P96311	4.39e-01
15	101	99.1	1426	10	P70298	7.51e-01
16	100	99.0	185	8	022194	9.79e-01
17	100	99.0	333	8	052486	9.79e-01
18	100	99.0	439	8	042421	9.79e-01
19	100	99.0	1048	10	063627	9.79e-01
20	100	99.0	1491	12	091718	9.79e-01

Result ID	Score	Query Match	Length	ID	Description	Pred. No.
21	99	8.9	556	9	046043	1.28e+00
22	99	8.9	642	8	023831	1.28e+00
23	98	8.8	351	8	039492	1.66e+00
24	98	8.8	1486	12	091717	1.66e+00
25	97	8.7	224	4	019031	2.15e+00
26	97	8.7	496	11	098457	2.15e+00
27	97	8.7	1337	2	015255	2.15e+00
28	97	8.7	4957	2	014687	2.15e+00
29	97	8.7	5262	2	014686	2.15e+00
30	96	8.6	356	9	006979	2.79e+00
31	96	8.6	381	2	015171	2.79e+00
32	96	8.6	400	12	013107	2.79e+00
33	96	8.6	605	2	015297	2.79e+00
34	96	8.6	1236	2	015398	2.79e+00
35	95	8.5	238	8	041883	3.61e+00
36	95	8.5	336	11	066620	3.61e+00
37	95	8.5	533	8	P93766	3.61e+00
38	95	8.5	556	2	015530	3.61e+00
39	95	8.5	665	9	048373	3.61e+00
40	95	8.5	732	2	013032	3.61e+00
41	95	8.5	1525	2	021920	3.61e+00
42	94	8.4	165	8	039352	4.67e+00
43	94	8.4	473	8	039620	4.67e+00
44	94	8.4	517	9	051619	4.67e+00
45	94	8.4	581	3	020517	4.67e+00

ALIGNMENTS

RESULT 1
 ID 092691; PRELIMINARY; PRT; 156 AA.
 AC 092691;
 DT 01-FEB-1997 (TREMBLREL, 02, CREATED)
 DT 01-FEB-1997 (TREMBLREL, 02, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL, 05, LAST ANNOTATION UPDATE)
 DE PRGI PROTEIN (DIF-2 PROTEIN).
 GN PRGI OR DIF-2.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 RN [1]
 RA SEQUENCE FROM N.A.
 RA TRAUZOLD A.;
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA PIETSCH A.; BUECHLER C.; ASLANIDIS C.; SCHMITZ G.;
 RA BIOCHEM. BIOPHYS. RES. COMMUN. 235:4-9(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA KONDRAITYEV A.D.; CHUNG K.N.; JUNG M.O.;
 RA CANCER RES. 56:1498-1502(1996).
 DR EMBL; Y14551; E333102;
 DR EMBL; X96438; E350480;
 SQ SEQUENCE 156 AA; 16928 MW; 46013510 CRC32;

Query Match 100.0%; Score 1114; DB 2; Length 156;
 Best Local Similarity 100.0%; Pred. No. 8.46e-174;
 Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query	156	Conservative	Mismatches	Indels	Gaps
1	MCHSRCHPTMTILQAPTPA	STIPGRSGPEITFDPLPEPAAPACRPSASRGHRK	60		
1	MCHSRCHPTMTILQAPTPA	STIPGRSGPEITFDPLPEPAAPACRPSASRGHRK	60		
61	RSRVLVPRVVRRLPVEEP	PARKLLFLTLTYFCOILAAE	60		
61	RSRVLVPRVVRRLPVEEP	PARKLLFLTLTYFCOILAAE	60		
121	APTVPSPVLEPNTSEPSD	VALDLSTFLQHPAAAF	156		
121	APTVPSPVLEPNTSEPSD	VALDLSTFLQHPAAAF	156		


```

RESULT 2
ID 085302 PRELIMINARY; PRT; 351 AA.
AC 085302;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE HOMOLOGUE OF RETROVIRAL PSEUDOPROTEASE.
OC VIRUS.
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; POXVIRIDAE; CHORDOPOXVIRINAE;
OC PARAPOXVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90266454.
RA FRASER M., HILL D.F., MERCER A.A., ROBINSON A.J.;
RL VIROLOGY 176:379-389(1990).
DR EMBL: M30023; G332564; -.
SQ SEQUENCE 351 AA; 39886 MW; 8EEF741B CRC32;

Query Match
Best Local Similarity 10.2%; Score 114; DB 11; Length 351;
Matches 24; Conservative 12; Mismatches 35; Indels 2; Gaps 2;

Db 236 QRRRRPRAAGARRGCGAPROQOPVQRAAAQORARQOROPPRVARRARAR 295
QY 15 QAPTPAPSTIPGRRGSGPEIFTFDLPPE-PAAPAGRPASRGHRSRVLYPRV-VR 72
Db 296 ROORAHORRRGR 308
QY 73 ROLPVEEPNPAKR 85

RESULT 3
ID P95306 PRELIMINARY; PRT; 115 AA.
AC P95306;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DE 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)
DE PREDICTED ORF.
OS MYCOBACTERIUM TUBERCULOSIS.
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RA TYAGI J.S.;
RL SUBMITTED (DEC-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RA VERMA A., DASGUPTA N., AGGARWAL A.N., PANDE J.N., TYAGI J.S.;
RL INDIAN J. BIOCHEM. BIOPHYS. 32:429-436(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RA TYAGI J.S.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: X63508; E282228; -.
SQ SEQUENCE 115 AA; 12106 MW; AF306566 CRC32;

Query Match
Best Local Similarity 10.1%; Score 112; DB 9; Length 115;
Matches 16; Conservative 23; Mismatches 22; Indels 2; Gaps 2;

Db 8 PSQALPPRPPTPAVPPPPAPPLDNNSSPLPTAPSPSPRRPNPAPARRRTAALR 67
QY 9 PTMTILOPPTAPSTIPGRRGSGPEIFTFDP-LPEPAAPAGRPSA-SRGHRSRVLY 66
Db 68 YRR 70
QY 67 YPR 69

RESULT 4
ID P78977 PRELIMINARY; PRT; 285 AA.

```

```

AC P78977;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CELL WALL PROTEIN PRECURSOR.
OS YAROMWIA LIPOLYTICA (CANDIDA LIPOLYTICA).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-CX39-74A; TISSUE-PINA240;
RX MEDLINE: 97127825.
RA RAMON A., GIL R., BURGAL M., SENTANDREU R., VALENTIN E.;
RL YEAST 12:1535-1548(1996).
DR EMBL: 281006; E274837; -.
KW SIGNAL.
FT SIGNAL.
FT CHAIN.
SQ SEQUENCE 285 AA; 30036 MW; 9AFCAB87 CRC32;

Query Match
Best Local Similarity 10.1%; Score 112; DB 1; Length 285;
Matches 43; Conservative 29; Mismatches 70; Indels 9; Gaps 9;

Db 122 ICHTITVTLSTEVIITPTVPQTTTPAVPRPTEVPEVKEPPTPEV-PGVKPEPTRCP 180
QY 1 MCHSRSCHPMTILOAPTPAD-STIPGRRGSGPEIFTFDLPPEPAAPAGRPASRGHR 59
Db 181 APKPEPEVPEVKEPPTPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEV 238
QY 60 KRSRRVLYPRVVRQPL-VEEENPAKRLLFLLLTVFCQIMAEQVPA-PLPEDAPNA 117
Db 239 LSLPPSRSSPSLPLSPLPTLPTSL-LS 268
QY 118 ASLAP-TP-VSPVLEPFTLSEPSDYALDS 146

RESULT 5
ID Q65553 PRELIMINARY; PRT; 3247 AA.
AC Q65553;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE UL36.
GN UL36.
OS BOVINE HERPESVIRUS TYPE 1.
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-COOPER;
RA SCHWYZER M., VLCEK C., LOWERY D.E., BELLO L.J., MEYER G., MISRA V.,
RA THIRY E., PACES V.;
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 1-179 FROM N.A.
RX STRAIN-COOPER;
RA SCHWYZER M., STYGER D., VOGT B., LOWERY D.E., SIMARD C.,
RA LABOISSIERE S., MISRA V., VLCEK C., PACES V.;
RL VET. MICROBIOL. 0:0-0(1996).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-COOPER;
RA SCHWYZER M.;
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE OF 1-179 FROM N.A.
RX STRAIN-JURA;
RA SCHWYZER M., STYGER D., VOGT B., LOWERY D.E., SIMARD C.,
RA LABOISSIERE S., MISRA V., VLCEK C., PACES V.;
RL VET. MICROBIOL. 53:67-77(1996).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN-JURA;
RA SCHWYZER M., VLCEK C., LOWERY D.E., BELLO L.J., MEYER G., MISRA V.;

```

```

FT CHAIN 8 228 PROLINE RICH PROTEIN.
SQ SEQUENCE 228 AA: 22480 MW: 2A2B03B9 CRC32;

Query Match 9.9%; Score 110; DB 8; Length 228;
Best Local Similarity 24.0%; Pred. No. 6,41e-02;
Matches 30; Conservative 36; Mismatches 56; Indels 3; Gaps 3;

Db 36 PPRPAPPTPTTQASPPVQS-SPPVQSSPPVQSSPPPASTPPVQSSPPVSAPP 94
16 APTAPPTITGPRRGSPELFTFDPLPEPAAAGRPSRGRRKRSRRVLYRVRRLQ 75

QY 95 PVQSSPPPTPLTPPPVOST-PPPASPPPASPPFSSPPATPPATPPATPPALPTPL 153
76 PVEPBNPAKRLLFLFLLITVFCQLMAEGVAPL-PREDAPNAASLAPTPSVLEPFL 134

Db 154 SSPPA 158
QY 135 TSEPS 139

RESULT 8 PRELIMINARY; PRT; 552 AA.
ID Q96343
AC Q96343;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DR 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DE 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE MYOOSINASE-BINDING PROTEIN RELATED PROTEIN (FRAGMENT).
OS BRASSICA NAPUS (RAPE).
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
OC CAPRATALES; CRUCIFERAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-20516 OF SVALOFS KARAT;
RA TALPALENSUD J., FALK A., EK B., RASK L.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U59446; G1655830; -.
FT NON TER 1
SQ SEQUENCE 552 AA: 59270 MW: 6861EF25 CRC32;

Query Match 9.7%; Score 108; DB 8; Length 552;
Best Local Similarity 43.5%; Pred. No. 1.12e-01;
Matches 20; Conservative 11; Mismatches 12; Indels 3; Gaps 3;

Db 235 LRTPAAPASPPAGPAPAPAGSGSH-APAPAPAPAGGGRAPAPAG 339
14 LQAPPAAPTIPGPRGSGPELFTFDPLPEPAAAP-A-G-RPSASRG 57

QY 14 LQAPPAAPTIPGPRGSGPELFTFDPLPEPAAAP-A-G-RPSASRG 57

RESULT 9 PRELIMINARY; PRT; 418 AA.
ID Q01662
AC Q01662;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DR 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DE 01-JUL-1997 (TREMBLREL. 04, LAST ANNOTATION UPDATE)
DE SIMILAR TO CUTICULAR COLLAGEN.
GN T8F2.6.
OS CAENORABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J.,
RA COULSON A., CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A.,
RA FULTON L., GARDNER A., GREEN P., HARKIN T., HILLIER L., JIER M.,
RA JOHNSON L., JONES M., KERSHAW J., KIRSTEN J., LANSTER N.,
RA LATEILLE P., LITCHING J., LLOYD C., MCMURRAY A., MORTIMORE B.,
RA SAUNDERS D., SHOWNKEEN R., PERCY C., RIEKEN L., ROOFA A.,
RA SAUDERS D., SHOWNKEEN R., SHALDON N., SMITH A., SONNHAMMER E.,
RA STADEN R., SUJSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,
RA VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.,

```



```

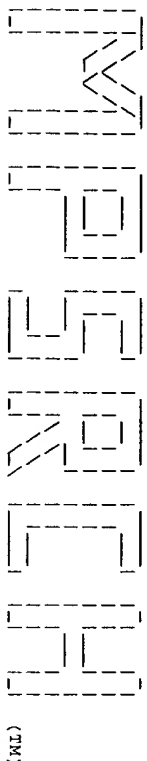
GN      CUTL2.
OS      MUS MUSCULUS (MOUSE).
OC      EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TERRAPODA; MAMMALIA.
OC      EUTHERIA; RODENTIA.
RN      [1]
RA      SEQUENCE FROM N.A.
RP      QUAGGIN S.E., VANSEN HEUYEL G.B., GOLDEN K., BOMMER R., IGARASHI P.;
RL      J. BIOL. CHEM. 271:22824-22834(1996).
DR      C-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC      EMBL; U45665; G1518932; -.
DR      MGD: MGI:107321; CUTL2.
DR      PROSITE; PS000027; HOMEBOX_1; 1.
KW      HOMEBOX; DNA-BINDING; NUCLEAR PROTEIN.
SQ      SEQUENCE 1426 AA; 154703 MW; 478F57CB CRC32;

Query Match          9.1%; Score 101; DB 10; Length 1426;
Best Local Similarity 34.8%; Pred. No. 7.51e-01;
Matches 16; Conservative 14; Mismatches 13; Indels 3; Gaps 3

Db      370 LGTEPPYPOLPPPGEDP-LSPPAQLIGP-SLGGDGRNFSL 413
      | : | | : | : | : | : | : | : | : | : | : | : |
QY      99 LMAEEGVAPRLPEDAPMAASLAPFVSPVLEPNTLSE-PSDYAL 143

Search completed: Fri Apr 17 13:39:48 1998
Job time : 47 secs.

```


 (TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1997 University of Edinburgh, U.K.
 Distribution Rights by Oxford Molecular Ltd

ch_n n.a. - n.a. database search, using Smith-Waterman algorithm

on: Thu Apr 23 05:25:50 1998; Maspar time 800.19 Seconds

Tabular output not generated. 1292.427 Million cell updates/sec

Title: >US-08-799-910-9

Description: (1-1228) from US08799910.seq

Perfect Score: 1228

N.A. Sequence: 1 ATGCTGCTACTCTCGACGCTG.....AAAAAAAAAACTCGAG 1228

Comp: TACACAGTGAAGCGCGAC.....TTTTTTTTTTTGAGCTC

Scoring table: TABLE default

Gap 6

Mmatch STD: Dbase 0; Query 0

Searched: 1126798 seqs, 421087984 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Database:

emb1_est 1:em_est1 2:em_est2 3:em_est3 4:em_est4 5:em_est5

6:em_est6 7:em_est8 8:em_gss

9:gb_est1 10:gb_est2 11:gb_est3 12:gb_est4 13:gb_est5

14:gb_est6 15:gb_est7 16:gb_est8 17:gb_est9 18:gb_est10

19:gb_est11 20:gb_est12 21:gb_gss 22:gb_est12 23:gb_est13

Statistics: Mean 11.559; Variance 3.326; scale 3.476

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description	Pred. No.
C 1	495	40.3	553	12	N32077	0.00e+00
C 2	450	36.6	459	14	W60982	0.00e+00
C 3	448	36.5	470	14	AA047094	0.00e+00
C 4	422	34.4	427	23	AA457705	0.00e+00
C 5	420	34.2	460	19	AA043722	0.00e+00
C 6	413	33.6	463	22	AA410666	0.00e+00
C 7	406	33.1	447	15	W52268	0.00e+00
C 8	383	31.2	481	19	AA034911	0.00e+00
C 9	382	31.2	413	15	W52269	0.00e+00
C 10	376	30.6	390	12	N29782	0.00e+00
C 11	373	30.4	443	15	W47587	0.00e+00
C 12	368	30.0	375	15	W47540	0.00e+00
C 13	351	28.1	355	14	AA047184	0.00e+00
C 14	345	28.1	405	9	T49532	0.00e+00
C 15	332	27.0	413	18	AA301628	0.00e+00

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT
1	N32077	553 bp mRNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 553)	1 (bases 1 to 553)	1 (bases 1 to 553)	1 (bases 1 to 553)
2	W60982	459 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 459)	1 (bases 1 to 459)	1 (bases 1 to 459)	1 (bases 1 to 459)
3	AA047094	470 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 470)	1 (bases 1 to 470)	1 (bases 1 to 470)	1 (bases 1 to 470)
4	AA457705	427 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 427)	1 (bases 1 to 427)	1 (bases 1 to 427)	1 (bases 1 to 427)
5	AA043722	460 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 460)	1 (bases 1 to 460)	1 (bases 1 to 460)	1 (bases 1 to 460)
6	AA410666	463 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 463)	1 (bases 1 to 463)	1 (bases 1 to 463)	1 (bases 1 to 463)
7	W52268	447 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 447)	1 (bases 1 to 447)	1 (bases 1 to 447)	1 (bases 1 to 447)
8	AA034911	481 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 481)	1 (bases 1 to 481)	1 (bases 1 to 481)	1 (bases 1 to 481)
9	W52269	413 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 413)	1 (bases 1 to 413)	1 (bases 1 to 413)	1 (bases 1 to 413)
10	N29782	390 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 390)	1 (bases 1 to 390)	1 (bases 1 to 390)	1 (bases 1 to 390)
11	W47587	443 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 443)	1 (bases 1 to 443)	1 (bases 1 to 443)	1 (bases 1 to 443)
12	W47540	375 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 375)	1 (bases 1 to 375)	1 (bases 1 to 375)	1 (bases 1 to 375)
13	AA047184	405 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 405)	1 (bases 1 to 405)	1 (bases 1 to 405)	1 (bases 1 to 405)
14	T49532	413 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 413)	1 (bases 1 to 413)	1 (bases 1 to 413)	1 (bases 1 to 413)
15	AA301628	413 bp cDNA	EST	10-JAN-1996	human clone-260175 primer-m13 -40 forward library-Soares placenta	Homo sapiens	1 (bases 1 to 413)	1 (bases 1 to 413)	1 (bases 1 to 413)	1 (bases 1 to 413)

ALIGNMENTS

human clone-260175 primer-m13 -40 forward library-Soares placenta
 809weeks 2NHP8T09W vector-p173D (Pharmacia) with a modified
 polylinker host-DH10B (ampicillin resistant) RseI+Not I
 RseI+2-Eco RI two placentae: one from 8 weeks and another from 9
 weeks post conception. 1st strand cDNA was primed with a Not I -
 15'-TGTTCACCACTGACGCGGAGCGGCGGATTTTCTTTT-3'),
 double-stranded cDNA was size selected, ligated to Eco RI adapters
 (Pharmacia), digested with Not I and cloned into the Not I and Eco
 RI sites of a modified pT73 vector (Pharmacia) library
 constructed by Bento Soares and M. Fatima Bonadio.
 Homo sapiens
 Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata;
 Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
 Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
 Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 553)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevas, E., Waterston, R., Williamson, A., Woldmann, P. and
 Wilson, R.
 The Washu-Merck EST Project
 Unpublished (1995)
 Contact: Wilson RK
 Washu-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 High quality sequence steps: 416

Source: IMAGE Consortium, LINTL
This clone is available royalty-free through LINTL ; contact the
IMAGE Consortium (info@image.lintl.gov) for further information.
Location/Qualifiers
Accession

FEATURES
source

```
/organism="Homo sapiens"  
/clone="260175"
```

	mrna	<1.	>553	
BASE COUNT	153	a	165	c
ORIGIN			106	g
			127	t
				2 others

Query Match	40.38;	Score 495;	DB 12;	Length 553;
Best Local Similarity	96.48;	Pred. No. 0.00e+00;		
Matches 528;	Conservative 0;	Mismatches 17;	Indels 3;	Gaps 2

Accession	Species	Gene	Position	Sequence
U00001	<i>Escherichia coli</i>	lacZ	100000	5' GAGCAGTAATCAATTTTATTTTGTGTCTCAGACAAATATCTAGCGATGCGATATCTC 64
U00001	<i>Escherichia coli</i>	lacZ	100000	CP 1204 GACAGTAAATTAATTTTATTTTGTGTCTCAGACAAATATCTAGCGATGCGATATCTC 1144
U00001	<i>Escherichia coli</i>	lacZ	100000	65 CGTGACAGCCCAACCAACCCCAACCCCTCTACCTCGACACCACCTAAAGGCACTTCAAG 124
U00001	<i>Escherichia coli</i>	lacZ	100000	1144 CGTGACAGCCCAACCAACCCCAACCCCTCTACCTCGACACCACCTAAAGGCACTTCAAG 108
U00001	<i>Escherichia coli</i>	lacZ	100000	125 AAGATGGAAGATCTCAGCGATCTCATTTCTTAATGTGTCCCGGAAGTCTCAGACAGTGA 18
U00001	<i>Escherichia coli</i>	lacZ	100000	CP 1084 AAGATGGAAGATCTCAGCGATCTCATTTCTTAATGTGTCCCGGAAGTCTCAGACAGTGA 102
U00001	<i>Escherichia coli</i>	lacZ	100000	185 CAGACGGAGTTGAGATGCTGGAGAGTGCAGTCACTCTCTAATCTAGACCCACACACAG 24
U00001	<i>Escherichia coli</i>	lacZ	100000	CP 1024 CAGACGGAGTTGAGATGCTGGAGAGTGCAGTCACTCTCTAATCTAGACCCACACACAG 96
U00001	<i>Escherichia coli</i>	lacZ	100000	245 ACTTCATCCCAAGCGCGAGAGTCTCTCCACCGACCGAGTCTCTCCATTTCTTCTCTACTT 30
U00001	<i>Escherichia coli</i>	lacZ	100000	CP 964 ACTTCATCCCAAGCGCGAGAGTCTCTCCACCGACCGAGTCTCTCCATTTCTTCTCTACTT 90
U00001	<i>Escherichia coli</i>	lacZ	100000	305 TGCCGCAATTCAGAGTGTCTCTGCTTCCACAGTCCGACAAAGTCTAATAATACCAGAG 36
U00001	<i>Escherichia coli</i>	lacZ	100000	CP 904 TGCCGCAATTCAGAGTGTCTCTGCTTCCACAGTCCGACAAAGTCTAATAATACCAGAG 84
U00001	<i>Escherichia coli</i>	lacZ	100000	365 ACCTCATTTACAGAGGAGGGAACATCTCACACCCCTTCATTAATTAATAATATTATAC 42
U00001	<i>Escherichia coli</i>	lacZ	100000	CP 844 ACCTCATTTACAGAGGAGGGAACATCTCACACCCCTTCATTAATTAATAATATTATAC 78
U00001	<i>Escherichia coli</i>	lacZ	100000	425 GTACACATCTCCATCACCTAGAGAGCGTACATTAATACATATATAATTAAGAGCAAT 48
U00001	<i>Escherichia coli</i>	lacZ	100000	CP 784 GTACACATCTCCATCACCTAGAGAGCGTACATTAATACATATATAATTAAGAGCAAT 72
U00001	<i>Escherichia coli</i>	lacZ	100000	485 AAG--GAATTAAATTAAACGACGCTTTTCTTCNACGAGGCGCTTAGCCAGCTGGAGT 54
U00001	<i>Escherichia coli</i>	lacZ	100000	724 AATAGAATAATTAATTAACGACGCTCT--CTTCCACACCGGCGCTTAGCCAGCTGGAGT 66
U00001	<i>Escherichia coli</i>	lacZ	100000	543 GCCCTGGT 550
U00001	<i>Escherichia coli</i>	lacZ	100000	CP 665 GCCCTGGT 658
U00001	<i>Escherichia coli</i>	lacZ	100000	RESULT 2
U00001	<i>Escherichia coli</i>	lacZ	100000	LOCUS W60982 459 bp mRNA EST 07-JUN-1996
U00001	<i>Escherichia coli</i>	lacZ	100000	DEFINITION zc98h09.s1 Pancreatic Islet Homo sapiens cDNA clone 339233 3'
U00001	<i>Escherichia coli</i>	lacZ	100000	ACCESSION W60982
U00001	<i>Escherichia coli</i>	lacZ	100000	KEYWORDS g1367741
U00001	<i>Escherichia coli</i>	lacZ	100000	EST.
U00001	<i>Escherichia coli</i>	lacZ	100000	SOURCE human.
U00001	<i>Escherichia coli</i>	lacZ	100000	ORGANISM Homo sapiens
U00001	<i>Escherichia coli</i>	lacZ	100000	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
U00001	<i>Escherichia coli</i>	lacZ	100000	Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo.
U00001	<i>Escherichia coli</i>	lacZ	100000	REFERENCE 1 (bases 1 to 459).
U00001	<i>Escherichia coli</i>	lacZ	100000	AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
U00001	<i>Escherichia coli</i>	lacZ	100000	Holman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
U00001	<i>Escherichia coli</i>	lacZ	100000	Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasakis, E.,
U00001	<i>Escherichia coli</i>	lacZ	100000	Waterston, R., Williamson, A., Woldmann, P. and Wilson, R.
U00001	<i>Escherichia coli</i>	lacZ	100000	TITLE WashU-Merck EST Project
U00001	<i>Escherichia coli</i>	lacZ	100000	JOURNAL Unpublished (1995)

[illegible]

QY	1038	ACTTCGGCGGACCATTTAGGAATGAGATCCGAGATCTTCATCTTCTTGAAGTGCGCT	1097
Db	361	TTAGGCTGGCTCCGAGGTAGAGAGGTTAGGGGTTTGTTGGGCTGTACAGGAGCAGTGTG	420
QY	1098	TTAGGCTGGCTCCGAGGTAGAGAGGTTGGGGGCTT-GGTGGGCTGTACAGGAGCAGTGTG	1156
Db	421	AGATCGCCCTAGTATGTTCTGTGACACAAATAAATGATTACTG	466
QY	1157	AGATCGCCCTAGTATGTTCTGTGACACAAATAAATGATTACTG	1202
RESULT	4	AA457705	421 bp
LOCUS		zx87c03.s1	Soares ovary tumor NDHOT Homo sapiens cDNA clone 810724
DEFINITION		3'	
ACCESSION		AA457705	
NID		92180425	
KEYWORDS		EST.	
SOURCE		human.	
ORGANISM		Homo sapiens	
		Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homindae;	
REFERENCE		1 (bases 1 to 427)	
AUTHORS		Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.	
TITLE		Wasnu-Merck EST Project 1997	
JOURNAL		Unpublished (1997)	
COMMENT		Contact: Wilson RK Wasnu-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available royalty-free through LNL; contact the IMAG Consortium (info@image.lln.gov) for further information. Seq primer: -14m13 fwd, RT from Amersham High quality sequence stop: 374.	
FEATURES		Location/Qualifiers	
source		1..427	
		/organism="Homo sapiens"	
		/note="Organ: ovary; Vector: pT73D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer (5' TGTTCACATCTGAGTGAGGAGCGCGCGCGGTTTCTTTTCTTTT 3')"	
		double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."	
		/db_xref="taxon:9606"	
		/db_xref="810724"	
		/clone.lib="Soares ovary tumor NDHOT"	
		/sex="Female"	
		/tissue_type="Ovarian tumor"	
		/lab_host="DH10B (ampicillin resistant)"	
		complement(1..>427)	
		/db_xref="GDB:6041404"	
BASE COUNT	118 a	139 c	78 g 92 t
ORIGIN			
Query Match	34.48;	Score 422;	DB 23; Length 427;
Best Local Similarity	99.84;	Pred. No. 0.00e+00;	
Matches	423;	Conservative 0;	Mismatches 1; Indels 0; Gaps 0;
Db	4	GACGTAATCAATTTTATTTGTTGTCACAGAACTAGTACGATCTGCAGATGCGCTC	63
QY	1204	GACGTAATCAATTTTATTTGTTGTCACAGAACTAGTACGATCTGCAGATGCGCTC	1145

		/db_xref="taxon:9606"	
		/clone_1487101"	
		/clone_1lib="Soares pregnant uterus NBHPU"	
		/sex="female"	
		/dev_stage="adult"	
		/lab_host="DH10B"	
		complement(<1..>460)	
		/db_xref="GDB:3760815"	
BASE COUNT		130 a	146 c
ORIGIN		83 g	98 t
		3 others	
Query Match 34.2%; Score 420; DB 19; Length 460;			
Best Local Similarity 98.3%; Pred No. 0.00e+00;			
Matches 451; Conservative 0; Mismatches 3; Indels 5; Gaps 5;			
Db 2 ACAGTAATCAATTTTATTTGTTGTTCCAGAAACATACATAAGGGATCTGCAGTGCCTC 61			
Cp 1203 ACAGTAATCAATTTTATTTGTTGTTCCAGAAACATACATAAGGGATCTGCAGTGCCTC 1144			
Db 62 GTGACAGCCCAACACCCCAACCCCTACTCTCGACGACCAACCCCTTAAAGGCACCTCAAGA 121			
Cp 1143 GTACAGACCCCAACACCCCAACCCCTACTCTCGACGACCAACCCCTTAAAGGCACCTCAAGA 1084			
Db 122 AGATGGAAGGATCTCAGGATCTCATTCCTTAATGGGTCCGCCGAAGTCTCACACAGTGA 181			
Cp 1083 AGATGGAAGGATCTCAGGATCTCATTCCTTAATGG-TCGCCGAAGTCTCACACAGTGA 1025			
Db 182 CAGACGAGTTGAGATGCTGAGAGATGCAATCCTCTTAACTTACGACCCACCCAGC 241			
Cp 1024 CAGACGAGTTGAGATGCTGAGAGATGCAATCCTCTTAACTTACGACCCACCCAGC 965			
Db 242 ACTTTCATCCAGCCGGAGAGTCTCTCCCAACCCGAGTCTCCCACTTCTTCTCTACTT 301			
Cp 964 ACTTTCATCCAGCCGGAGAGTCTCTCCCAACCCGAGTCTCCCACTTCTTCTCTACTT 905			
Db 302 TGGCGCAGTTCCAGAGTGTCTCTGCTTCCAGCAGTCCCAACAACTCAATTAATACCAAG 361			
Cp 904 TGGCGCAGTTCCAGAGTGTCTCTGCTTCCAGCAGTCCCAACAACTCAATTAATACCAAG 845			
Db 422 ACCTGATTTTACAGCAGGGGGAACATCTCAACCCCTTGGCATTAAGTTAAATTAATATT 421			
Cp 844 ACCTGATTTTACAGCAGGGGGAACATCTCAACCCCTTG-CATTA-GTTAAATTAATTAATT 787			
Db 422 ACCTGATTTTACAGCAGGGGGAACATCTCAACCCCTTGGCATTAAGTTAAATTAATATT 460			
Cp 786 ACCTGATTTTACAGCAGGGGGAACATCTCAACCCCTTGGCATTAAGTTAAATTAATATT 750			
RESULT	6	463 bp	EST
LOCUS	AA410666	463 bp	18-MAY-1997
DEFINITION	5' similar to SW:GL96_MOUSE P46694 IMMEDIATE EARLY PROTEIN GL196.		
ACCESSION	AA410666		
NID	92069789		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	mitochondrial eukaryotes; Metazoa; Chordata;		
	Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae;		
	Homo.		
REFERENCE	1 (bases 1 to 463)		
AUTHORS	Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gaisel, G., Jost, S.,		
	Kucaba, T., Lacy, M., Le, N., Lennon, G., Marr, M., Martin, J.,		
	Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,		
	White, Y., Wylie, T., Waterston, R. and Wilson, R.		
TITLE	Washu-Merck EST Project 1997		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Wilson RK		
	Washu-Merck EST Project		
	Washington University School of Medicine		

Db	300	GGTGGCGCGGGCCCGCTCCCTCAGAGAGAGCGCCCTTAACGCGCATCCCTGGGCGCCAC	359
Oy	310	GGTGTGGCGGGCGCCCTCCCTCAGAGAGAGCGCCCTTAACGCGCATCCCTGGGCGCCAC	368
Db	360	CCCTGTGTCCCCCGCTCCCTCAGAGAGCGCCCTTAACCTGACTTCGAGAGCCCTCGAGTACGCTCT	419
Oy	369	CCCTGTGTCCCCCGCTCCCTCAGAGAGCGCCCTTAACCTGACTTCGAGAGCCCTCGAGTACGCTCT	428
Db	420	GGACCTCAGCACTTTCCTCCAGCAACAC	447
Oy	429	GGACCTCAGCACTTTCCTCCAGCAACAC	456

RESULT	8	AA034911	481 bp	mRNA	EST	10-MAY-1997
LOCUS		2625501.1	Soares pregnant uterus	NBHPU	Homo sapiens	CDNA clone
DEFINITION		471552.3				
ACCESSION		AA034911				
NID		01506874				
KEYWORDS		EST.				
SOURCE		human.				
ORGANISM		Homo sapiens				
		Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae;				
REFERENCE		1 (bases 1 to 481)				
AUTHORS		Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.				
TITLE		Mashu-Merck EST project				
JOURNAL		Unpublished (1995)				
COMMENT		Contact: Wilson RK Mashu-Merck EST project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 1245 Std Error: 0.00 Seq primer: -40M13 fwd. from Amersham High quality sequence stop: 420. Location/Qualifiers 1. 481 /organism="Homo sapiens" /note="Organ: uterus; Vector: pT773-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand CDNA was primed with a Not I - oligo(dT) primer [5', AACTGGAAGAATTCGCGCGCCCTTTTCTTTTCTTTTCTTTT 3'] double-stranded CDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT773 vector. Library went through one round of normalization. Library constructed by M. Fatiima Bonaldo." /db_xref="taxon:9606" /clone="471552" /clone_11b="Soares pregnant uterus NBHPU" /sex="female" /dev_stage="adult" /lab_host="DH10B" complement(<1. >481) /db_xref="GDB:3757554"				

BASE COUNT	140 a	139 c	85 g	105 t	12 others
ORIGIN					
Query Match	32.7%	Score 402;	DB 19;	Length 481;	
Best Local Similarity	95.4%;	Pred. 0.00e+00;			
Matches	459;	Conservative	0;	Mismatches 13;	Indels 9;
				Gaps	7;

1 GACAGTAAATCAATTTATTTGTGTTCACAGAACTACTAGGCGATCTCGACAGTCGCTC 60

CP	1204	GAGAGTAATCAATTTTATTTTGTGTGCAGAGACATCTAGGGGATCTGCAGACGTGCTC	1145
Db	61	CGTAGACGCCCAACCAACCCCAACCCCTTACCTTGACAGCCACCCCTAAAGGCGAATTCAAG	120
Cp	1144	CGTAGACGCCCAACCAACCCCAACCCCTTACCTTGACAGCCACCCCTAAAGGCGAATTCAAG	1085
Db	121	AAAGTAGAAGGATCTCAGCGATCTCTATTCTCTAATNGTGTCCGCCAAGTCTCAGCAAGTAA	180
Cp	1084	AAAGTAGAAGGATCTCAGCGATCTCTATTCTCTAATNGTGTCCGCCAAGTCTCAGCAAGTAA	1025
Db	181	CAGACGAGTGGATGCTGTGGAGATGAGTACAGTCCACTCTTAACTTGTAGACCCACGACG	240
Cp	1024	CAGACGAGTGGATGCTGTGGAGATGAGTACAGTCCACTCTTAACTTGTAGACCCACGACG	965
Db	241	ACTTCNTACNCNANGAGGAGAGCTCCTCCGCCANCGANGAGTCCCTCCATTCTCTCC	300
Cp	964	ACTTCNTACNCNANGAGGAGAGCTCCTCCGCCANCGANGAGTCCCTCCATTCTCTCC	910
Db	301	TACTTTGCGGNAAGTTCAGAGTGTCTCTCTCTTCCACAGTCCACACAAAGCTCAATTAATAC	360
Cp	909	TACTTTGCGGNAAGTTCAGAGTGTCTCTCTCTTCCACAGTCCACACAAAGCTCAATTAATAC	851
Db	361	CAGAGACCGGATTTTACAGCAGAGGGGAACANTGGAANCGTTGGCATTAAGTTAAAT	420
Cp	850	CAGAGACCGGATTTTACAGCAGAGGGGAACANTGGAANCGTTGGCATTAAGTTAAAT	794
Db	421	AAATATTACGATCATCTCTCATCACCACGAGAGAGCTAGATTAATCATATTAATTAATN	480
Cp	793	AAATATTACGATCATCTCTCATCACCACGAGAGAGCTAGATTAATCATATTAATTAATN	734
Db	481	A 481	
Cp	733	A 733	
RESULT	9		
LOCUS	W52269	413 bp	mRNA
DEFINITION	zc46a10.s1 Soares senescent fibroblasts NBHSF Homo sapiens CDNA		
ACCESSION	U52269		
NID	91349381		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotic mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 413) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Treviski,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.		
TITLE	The WashU-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 3171 Std Error: 0.00 Seq primer: mob.REGA+ET. Location/Qualifiers 1. 413 /organism="Homo sapiens" /note="vector: p773D (Pharmacia) with a modified polylinier V-type; phagemid; Site_1: Not I; Site_2: Eco RI; TGTTACCAATCTGAAGTGGAGCGGCCGCAATTTTTTTTTTTTTTTTTT		
FEATURES			
source			

3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo.

/db.xref="taxon:9606"
/clone="325338"
/tissue_type="senescent fibroblasts NbHSF"
/lab_host="DH10B (ampicillin resistant)"
complement(<1..>413)

BASE COUNT 110 a 136 c 77 g 89 t 1 others
ORIGIN

Query Match 31.2%; Score 383; DB 15; Length 413;
Best Local Similarity 98.8%; Pred. No. 0.00e+00;
Matches 408; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

1 GACAGTAATCAATTTATTTGTTGTCACAGAACACTAGCGATCTGACAGTGCCTC 60
|||||
CP 1204 GACAGTAATCAATTTATTTGTTGTCACAGAACACTAGCGATCTGACAGTGCCTC 1145
61 CGTGACAGCCCAACCAACCCCTCTACCTCGACGACCCCTAAAGCGACTTCAAG 120
|||||
CP 1144 CGTGACAGCCCAACCAACCCCTCTACCTCGACGACCCCTAAAGCGACTTCAAG 1085
121 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGTCGCGGAGCTTCACACAGTGA 180
|||||
CP 1084 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGTCGCGGAGCTTCACACAGTGA 1025
181 CAGACGAGTTAGATGTTGATGAGATGACATCACTCTAACTTAAGCCACACACCA 240
|||||
CP 1024 CAGACGAGTTAGATGTTGATGAGATGACATCACTCTAACTTAAGCCACACCA 966
241 GACTTCATCCAGCCGCGGAGCTCTCCACCCGAGTCTCTCCCATTTCTTCTCTACT 300
|||||
CP 965 GACTTCATCCAGCCGCGGAGCTCTCCACCCGAGTCTCTCCCATTTCTTCTCTACT 906
301 TTGCGCGAGTTCCAGATGTTCTGCTTCCACCAAGTCCACAAAGCTCAATTAATTCGA 360
|||||
CP 905 TT-GCGCGAGTTCCAGATGTTCTGCTTCCACCAAGTCCACAAAGCTCAATTAAT-ACCA 848
361 GAGACCTGATTTACAGACAGGAGGAACTCTACACCTCTGATTAAGTTAA 413
|||||
CP 847 GAGACCTGATTTACAGACAGGAGGAACTCTACACCTCTGATTAAGTTAA 796

RESULT 10
LOCUS N29782 390 bp mRNA EST 05-JAN-1996
DEFINITION YW91h09.s1 Homo sapiens cDNA clone 259649 3'
ACCESSION N29782
NID 91148302
KEYWORDS
SOURCE

human clone-259649 primer-m13 -40 forward library-Soares placenta 8to9weeks 2Nbp8to9w vector-pT73D (Pharmacia) with a modified polylinker host-DH10B (ampicillin resistant) Rstel-Not I Rstel2-Eco RI two placentae: one from 8 weeks and another from 9 weeks post conception. 1st strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAGAGGAGGCGGCGGATTTTCTTTTCTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;

Sarcopterygia; Choanata; Tetrapoda; Amniota; Mammalia; Theria;

Eutheria; Archonta; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 390)

REFERENCE

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washu-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Location/Qualifiers
1..390
/organism="Homo sapiens"
/clone="259649"

BASE COUNT 103 a 129 c 75 g 83 t
ORIGIN

Query Match 30.6%; Score 376; DB 12; Length 390;
Best Local Similarity 98.7%; Pred. No. 0.00e+00;
Matches 381; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

5 GACAGTAATCAATTTATTTGTTGTCACAGAACACTAGCGATCTGATGCGCTC 64
|||||
CP 1204 GACAGTAATCAATTTATTTGTTGTCACAGAACACTAGCGATCTGATGCGCTC 1145
65 CGTGACAGCCCAACCAACCCCTCTACCTCGACGACCCCTAAAGCGACTTCAAG 124
|||||
CP 1144 CGTGACAGCCCAACCAACCCCTCTACCTCGACGACCCCTAAAGCGACTTCAAG 1085
125 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGTCGCGGAGCTTCACACAGTGA 184
|||||
CP 1084 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGTCGCGGAGCTTCACACAGTGA 1025
185 CAGACGAGTTAGATGTTGATGAGATGACATCACTCTAACTTAAGCCACACCA 244
|||||
CP 1024 CAGACGAGTTAGATGTTGATGAGATGACATCACTCTAACTTAAGCCACACCA 965
245 ACTTCATCCAGCCGCGGAGCTCTCCACCCGAGTCTCTCCCATTTCTTCTCTACT 304
|||||
CP 964 ACTTCATCCAGCCGCGGAGCTCTCCACCCGAGTCTCTCCCATTTCTTCTCTACT 905
305 TGCCGAGTTCCAGGATGCTCTCTCCACAGTCCACAAAGCTCAATTAATTCAGAG 364
|||||
CP 904 TGCCGAGTTCCAGGATGCTCTCTCTCCACAGTCCACAAAGCTCAATTAATTCAGAG 845
365 ACCTGCATTTACAGCAGGAGGAACAT 390
|||||
CP 844 ACCTGCATTTACAGCAGGAGGAACAT 819

RESULT 11
LOCUS W47587 443 bp mRNA EST 11-OCT-1996
DEFINITION ZC35B02.t1 Soares senescent fibroblasts NbHSF Homo sapiens cDNA clone 32467 5' similar to SW:GI96_MOUSE P46594 IMMEDIATE EARLY PROTEIN G1Y96. [1] .
ACCESSION W47587
NID 91332247
KEYWORDS
SOURCE

human.

Eukaryota; Eukaryotes; Eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 443)

REFERENCE

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

CP 1024 CAGACGGAGTTGAGATGCTGGAGATGAGTACCTCTCTAAACTACGACCCACCAACG 965

DB 244 ACTTCATCCAGCCGGAGAGTCTCTCCGCCAGTCTCTCCCACTTTCTCTCTACTT 303

CP 964 ACTTCATCCAGCCGGAGAGTCTCTCCGCCAGTCTCTCCCACTTTCTCTCTACTT 905

DB 304 TGCCGCAAGTTCAGAGTGTCTCTCTCTCCAGCAGTCCACAAAGCTCAATTAATACCAAG 363

CP 904 TGCCGCAAGTTCAGAGTGTCTCTCTCTCCAGCAGTCCACAAAGCTCAATTAATACCAAG 845

DB 364 ACCTGATTTA 374

CP 844 ACCTGATTTA 834

RESULT 13
LOCUS AA047184 355 bp mRNA EST 06-SEP-1996
DEFINITION ZK42402.s1 Soares pregnant uterus NBHPU Homo sapiens cDNA clone 48522.3'

ACCESSION AA047184
NID 91525150
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 355)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 324.
FEATURES
Location/Qualifiers
1..355
/organism="Homo sapiens"
/note="Organ: uterus; Vector: pT73-Pac; Site: 1: Not I; Site: 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15', AACGGAAGATTCGGCGCGCCCTTTTCTTTTCTTTT 3', AACGGAAGATTCGGCGCGCCCTTTTCTTTTCTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="48522"
/clone_lib="Soares pregnant uterus NBHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
complement(1..>355)
BASE COUNT 91 a 121 c 66 g 76 t 1 others
ORIGIN
Query Match 28.6%; Score 351; DB 14; Length 355;
Best Local Similarity 99.7%; Pred. No. 0.00e+00;
Matches 351; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CP 1204 GACAGTAATCAATTTATTGTGTTCAGAAATATCTAGGAGATCTGACAGTGGCTC 1145

DB 64 CGTGACAGCCACCAACCCCAACCTCTACCTGCGAGCCACCTTAAGGAGACTTCAAG 123

CP 1144 CGTGACAGCCACCAACCCCAACCTCTACCTGCGAGCCACCTTAAGGAGACTTCAAG 1085

DB 124 AAGATGGAAGATCTACAGGATCTCATCTCTAATGATCGCCGGAAGTCTCACACAGTGA 183

CP 1084 AAGATGGAAGATCTACAGGATCTCATCTCTAATGATCGCCGGAAGTCTCACACAGTGA 1025

DB 184 CAGACGGAGTTGAGATGCTGGAGATGAGTACCTCTCTAAACTACGACCCACCAACG 243

CP 1024 CAGACGGAGTTGAGATGCTGGAGATGAGTACCTCTCTAAACTACGACCCACCAACG 965

DB 244 ACTTCATCCAGCCGGAGAGTCTCTCCGCCAGTCTCTCCCACTTTCTCTCTACTT 303

CP 964 ACTTCATCCAGCCGGAGAGTCTCTCCGCCAGTCTCTCCCACTTTCTCTCTACTT 905

DB 304 TGCCGCAAGTTCAGAGTGTCTCTCTCTCCAGCAGTCCACAAAGCTCAATTAAT 355

CP 904 TGCCGCAAGTTCAGAGTGTCTCTCTCTCCAGCAGTCCACAAAGCTCAATTAAT 853

RESULT 14
LOCUS T49532 405 bp mRNA EST 08-FEB-1995
DEFINITION ya76f12.s1 Homo sapiens cDNA clone 67631.3'

ACCESSION T49532
NID 9651392
KEYWORDS EST.
SOURCE human clone-67631 library-Stratagene placenta (#937225)
vector-pluscript SK- host-SOLR cells (kanamycin resistant)
primer-21m3 Raitel-Ecor1 Raitel2-XhoI Placental tissue from a Caucasian male. Cloned unidirectionally. Primer: Oligo dT. Average insert size: 1.2 kb; Uni-ZAP XR Vector; 5' adaptor sequence: 5'-GAATTCGACGACGAG-3', 3' adaptor sequence: 5'-CTCGAGTCTTTTCTTTTCTTTT-3'.
ORGANISM Homo sapiens
Eukaryotae; Metazoa; Chordata; Vertebrata; Hominiidae; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 405)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Other ESTs: ya76f12.r1
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES
Location/Qualifiers
1..405
/organism="Homo sapiens"
/clone="67631"
BASE COUNT 108 a 132 c 81 g 82 t 2 others
ORIGIN
Query Match 28.1%; Score 345; DB 9; Length 405;
Best Local Similarity 96.3%; Pred. No. 0.00e+00;
Matches 388; Conservative 0; Mismatches 9; Indels 6; Gaps 6;

DB 4 GACAGTAATCAATTTATTGTGTTCAGAAATATCTAGGAGATCTGACAGTGGCTC 63


```

Query Match      8.8%  Score 98; DB 1; Length 475;
Best Local Similarity 28.6%  Pred. No. 6,056-01;
Matches 18; Conservative 19; Mismatches 24; Indels 2; Gaps 2

Db      358  LICPILFDKVTYTDMLAMEE-PFGVPLPIIRYTSVEEAIEISNK-SEYGLQASIFINDFP 415
Oy      94  VECQILMAEEGVAPALPEDAPNMAASLAPFTVSPVLEPFNLTSPESDYALDLSIFLQDHP 153
Db      416  RAF 418
Oy      154  AAF 156

RESULT 15
ID  ALAB_HUMAN  STANDARD:  PRT:  519 AA.
AC  P35368:
DT  01-JUN-1994 (REL. 29, CREATED)
DT  01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT  01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE  ALPHA-1B ADRENERGIC RECEPTOR.
GN  ADRA1B.
    HOMO SAPIENS (HUMAN).
    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
    EUHARERIA; PRIMATES.
    [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE: 93016158.
RA  RAAMARO C.S., DENKER J.M., PEREZ D.M., GAIVIN R.J., RIEK R.P.,
RA  GRAHAM R.M.;
RL  J. BIOL. CHEM. 267:21936-21945(1992).
RP  [2]
RP  SEQUENCE FROM N.A.
RX  TISSUE-BRAIN:
RX  MEDLINE: 94239386.
RA  FORRAY C., BARD J.A., WETZEL J.M., CHIU G., SHAPIRO E., TANG R.,
RA  LEORAY H., HARTIG P.R., WEINSHANK R.L., BRANCHER T.A.,
RA  GLUCHOMSKI C.;
RL  MOL. PHARMACOL. 45:703-708(1994).
RN  [3]
RP  SEQUENCE FROM N.A.
RX  MEDLINE: 95114877.
RA  SCHWAIN D.A., JOHNSTON G.I., PAGE S.O., MOSLEY M.J., WILSON K.H.,
RA  WOMAN N.P., CAMPBELL S., FIDOCK M.D., FURNESS L.M.,
RA  PARKY SMITH D.J., PETER B., BAILEY D.S.;
RL  J. PHARMACOL. EXP. THER. 272:134-142(1995).
CC  -1- FUNCTION: THIS ALPHA-ADRENERGIC RECEPTOR MEDIATES ITS ACTION BY
CC  ASSOCIATION WITH G PROTEINS THAT ACTIVATE A PHOSPHATIDYLINOSITOL-
CC  CALCIUM SECOND MESSENGER SYSTEM.
CC  -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC  -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
EMBL: M99589; -; NOT_ANNOTATED_CDS.
EMBL: M99590; -; NOT_ANNOTATED_CDS.
EMBL: U03865; G494983; -.
PIR: L31773; G666891; -.
PIR: A45121; A45121.
GCRDB: GCR_0530; -.
DR  MIM: 104220; -.
DR  PROSITE: PS00237; G-PROTEIN_RECEPTOR. 1.
KW  G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW  MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT  DOMAIN 1 45
FT  TRANSMEM 46 70
FT  DOMAIN 71 83
FT  TRANSMEM 84 105
FT  DOMAIN 106 115
FT  TRANSMEM 116 141
FT  DOMAIN 142 161
FT  TRANSMEM 162 182
FT  DOMAIN 183 201
FT  TRANSMEM 202 224
FT  DOMAIN 225 295
FT  TRANSMEM 296 319
FT  DOMAIN 320 326
FT  EXTRACELLULAR (POTENTIAL).

```

FT	TRANSMEM	327	340	7 (POTENTIAL).
FT	DOMAIN	341	519	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	370	379	POLY-ARG.
FT	CARBOHYD	10	10	POTENTIAL.
FT	CARBOHYD	24	24	POTENTIAL.
FT	CARBOHYD	29	29	POTENTIAL.
FT	CARBOHYD	34	34	POTENTIAL.
FT	DISULFID	118	195	BY SIMILARITY.
FT	LIPID	365	365	PALMITATE (POTENTIAL).
FT	CONFLICT	379	379	MISSING (IN REF. 1).
FT	CONFLICT	497	500	AAAD -> PRH (IN REF. 1).
FT	CONFLICT	370	370	R -> RG (IN REF. 2).
SO	SEQUENCE	519 AA;	56778 MM;	D0868586 CAC32.
Query Match				
Best Local Similarity		41.3%;	Score 98;	DB.1; Length 519;
Matches 19;		Conservative	9;	Mismatches 15; Indels 3; Gaps 3.
Db	451	ALTSIPAPEPPGRG-RHDSGP-LTFEKLLEPEPS-EGTDSGASNG	493	
Oy	12	TTLAPPTAPPTIPGRGSGPELFTFDPLPEAPAAAGRSASRG	57	

Search completed: Fri Apr 17 13:38:42 1998
Job time : 10 secs.

GN CEBPD OR CELF.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC EUHERIA; RODENTIA.
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 91332085.
RA KAGEYAMA R., SASAI Y., NAKANISHI S.;
RL J. BIOL. CHEM. 266:15525-15531(1991).
CC -1- FUNCTION: C/EBP IS A DNA-BINDING PROTEIN THAT RECOGNIZES TWO
CC DIFFERENT MOTIFS: THE CCAAT HOMOLOGY COMMON TO MANY PROMOTERS AND
CC THE ENHANCED CORE HOMOLOGY COMMON TO MANY ENHANCERS.
CC -1- SUBUNIT: BINDS DNA AS A DIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: UNBIOLOGICALLY EXPRESSED.
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS. STRONG. TO OTHER C/EBP
CC PROTEINS.
DR EMBL; M65149; G203412; -
DR PIR; B39429; B39429.
DR TRANSFAC; T00109; -
DR TRANSCRIPTION REGULATION; ACTIVATOR: DNA-BINDING; NUCLEAR PROTEIN.
CC DNA_BIND 197 215 BASIC MOTIF.
DOMAIN 226 254 LEUCINE-ZIPPER.
SQ SEQUENCE 268 AA; 28600 MW; 760E79C0 CRC32;
Query Match 9.1%; Score 101; DB 1; Length 268;
Best Local Similarity 36.7%; Pred. No. 2.57e-01;
Matches 22; Conservative 10; Mismatches 25; Indels 3; Gaps 3;
Db 142 CAQTVSLAAMAPPTSPSPSPSLAP-GPVREKAGKGRPPGSPPEYRQRER 200
Oy 7 CHTPTIIQAPF-PAPSTIPGRGS-GPEITFDLPDEPAAPAGRSASGCHKRSR 64
RESULT 9
ID A1AB-CA_NFA STANDARD; PRT; 417 AA.
AC P1615;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE ALPHA-1B ADRENERGIC RECEPTOR (FRAGMENT).
GN ADRA1B OR RDC5.
OS CANIS FAMILIARIS (DOG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC EUHERIA; CARNIVORA.
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE-THYROID;
RX MEDLINE; 89242119.
RA LIBERT F., PARMENTIER M., LEFORT A., DINSART C., VAN SANDE J.,
RL MAENHAUT C., SIMONS M.-J., DUMONT J.E., VASSART C.;
SCIENCE 244:569-572(1989).
CC -1- FUNCTION: THIS ALPHA-ADRENERGIC RECEPTOR MEDIATES ITS ACTION BY
CC ASSOCIATION WITH G PROTEINS THAT ACTIVATE A PHOSPHATIDYLINOSITOL-
CC CALCIUM SECOND MESSENGER SYSTEM.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; X14050; G833784; -
DR PIR; E30341; G30341.
DR GCRDB; GCR_0017; -
DR PROSITE; PS00237; G-PROTEIN RECEPTOR. 1.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.
FT NON_TER 1 1
FT TRANSMEM 1 8
FT DOMAIN 9 20 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 21 44 3 (POTENTIAL).
FT DOMAIN 45 64 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 65 85 4 (POTENTIAL).
FT DOMAIN 86 104 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 105 127 5 (POTENTIAL).
FT DOMAIN 128 198 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 199 222 6 (POTENTIAL).
FT DOMAIN 223 229 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 230 243 7 (POTENTIAL).
FT DOMAIN 244 417 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 271 281 POLY-ARG.
FT DISULFID 21 98 BY SIMILARITY.
FT LIPID 268 268 PALMITATE (POTENTIAL).
SQ SEQUENCE 417 AA; 45891 MW; 8008E9B3 CRC32;
Query Match 9.1%; Score 101; DB 1; Length 417;
Best Local Similarity 43.6%; Pred. No. 2.57e-01;
Matches 17; Conservative 10; Mismatches 10; Indels 2; Gaps 2;
Db 353 ALLSPAPOPGRRG-RDSCP-LFTFRLAERSSPAG 389
Oy 12 TILQPTPAPSTIPGRSGPEITFDLPDEPAAPAG 50
RESULT 10
ID QALF-NEUCR STANDARD; PRT; 816 AA.
AC P1638;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE QUINIC ACID UTILIZATION ACTIVATOR.
GN QA-1F.
OS NEUROSPORA CRASSA.
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; PYRENOMYCETES; SORDARIALES.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-74-OR23-1A;
RX MEDLINE; 89293848.
RA GEVER R.F., HUET L., BAUM J.A., TYLER B.M., PATEL V.B.,
RA RUTLEDGE B.J., CASE M.E., GILES N.H.;
RL J. MOL. BIOL. 207:15-34(1989).
CC -1- FUNCTION: TRANSCRIPTION ACTIVATION OF GENES FOR ENZYMES AND
CC PROTEINS OF QUINATE METABOLISM BY BINDING TO A 16 BASE-PAIR
CC SEQUENCE (CONSENSUS GGTARARYTTAYCC) IN FRONT OF EACH QA GENE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC CLUSTER DOMAIN.
DR EMBL; X14603; G3068; -
DR PIR; S04256; S04256.
DR PIR; F31277; F31277.
DR TRANSFAC; T00709; -
DR PROSITE; PS00463; ZN2-CYS6-FUNGAL_1; 1.
DR PROSITE; PS00463; ZN2-CYS6-FUNGAL_2; 1.
KW TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING; NUCLEAR PROTEIN;
KW ZINC; METAL-BINDING; QUINATE METABOLISM.
FT DNA_BIND 76 103 ZN(2)-CYS(6), FUNGAL-TYPE.
FT DOMAIN 773 793 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 816 AA; 88946 MW; E59EACAF CRC32;
Query Match 9.1%; Score 101; DB 1; Length 816;
Best Local Similarity 33.3%; Pred. No. 2.57e-01;
Matches 16; Conservative 13; Mismatches 18; Indels 1; Gaps 1;
Db 408 VVRQLPDDDRPRRRALLVACFYDVTYSMRHNPALKPDLD 455
Oy 70 VVRQLPDDDRPRRRALLVACFYDVTYSMRHNPALKPDLD 116
RESULT 11
ID GSPC-ERMCH STANDARD; PRT; 272 AA.
AC P31696;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE GENERAL SECRETION PATHWAY PROTEIN C (PECTIC ENZYMES SECRETION PROTEIN
DE OUTC).
GN OUTC.
OS ERWINIA CHRYSANTHEMI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
 RA FULTON L., GARNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,
 RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
 RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
 RA SUSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
 RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
 RA WOHLDMAN P.,
 RL NATURE 368:32-38(1994).
 DR EMBL: L16621; G289783; -
 DR PIR: S44920; S44920.
 DR MORPEP: 2K688.5; CE00463.
 DR PROSITE: P55053; UNIQUTIN_2; UNKNOWN_1.
 KW HYPOTHETICAL PROTEIN.
 -- SEQUENCE 1799 AA; 202641 MW; 3EF7DDB2 CRC32;

Query Match 9.8%; Score 109; DB 1; Length 1799;
 Best Local Similarity 23.7%; Pred. No. 2,43e-02;

Matches 31; Conservative 30; Mismatches 65; Indels 5; Gaps 5;

DB 844 FLENRGRIPTSSAPSTSENPSPSPFNS-EDAADIRAGRLPLGTRPNRTVRETVHPAA 902
 QY 13 ILQAPRPAPSTIGPRGSGPLFTDPLEPPAAPAGR-PSMSRGHRKRSKRVLYPRV 71
 DB 903 ARAESPNSHLTFTATTHFPAGF-PLMASSNVSTSGPGWPRIQVSPPTTGL 961
 QY 72 RR-OLVEEPNPAKRLFLTLTIVFCQILMAEGVPALP-PEDAPNAASLAPTPVSL 129
 DB 962 FEEDLGSSDQ 972
 QY 130 EPNLTSEPSD 140

RESULT 6 STANDARD; PRT; 1206 AA.
 AC 005859;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE FORMIN 4 (LIMB DEFORMITY PROTEIN).
 GN LD.
 OS MUS MUSCULUS (MOUSE).
 KW EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 -- EUTHERIA; RODENTIA.
 -- [1]
 -- SEQUENCE FROM N.A.
 -- TISSUE-EMBRYO;
 -- MEDLINE: 92112033.
 -- RA GRUSBY-JACKSON L., KUO A., LEDER P.;
 -- RL GENES DEV. 6:29-37(1992).
 -- CC -1- FUNCTION: IS IMPORTANT IN THE MORPHOGENESIS OF LIMB AND MAY HAVE A
 -- FUNCTION IN DIFFERENTIATED CELLS OR BE INVOLVED IN MAINTAINING
 -- SPECIFIC DIFFERENTIATED STATES.
 -- CC -1- TISSUE SPECIFICITY: IT IS FOUND THROUGHOUT THE EMBRYO BUT
 -- HAS A FUNCTIONAL ROLE ONLY IN THE KIDNEY AND LIMB.
 -- CC -1- DEVELOPMENTAL STAGE: THIS IS THE ISOFORM FOUND IN THE APICAL
 -- ECTODERMAL RIDGE AND THE MESENCHYMAL COMPARTMENT OF THE DEVELOPING
 -- LIMB BUD.
 -- CC -1- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE
 -- PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A VARIATION IN
 -- SPLICING IS SEEN AMONG DIFFERENT TISSUES AND DIFFERENT SIDE
 -- TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE. THIS IS ISOFORM 4 AND IS
 -- DIFFERENT IN ITS N-TERMINAL TO THE OTHER MOUSE ISOFORMS SO FAR
 -- DETERMINED.
 -- CC EMBL: X62379; G51553; -
 -- DR PIR: S24407; S24407.
 -- DR HSSP: P19999; ICLG.
 -- KW NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING.
 -- FT DOMAIN 635 638 POLY-SER.
 -- FT DOMAIN 644 744 PRO-RICH.

FT DOMAIN 751 755 POLY-SER.
 SQ SEQUENCE 1206 AA; 133464 MW; 6D70C261 CRC32;

Query Match 9.2%; Score 103; DB 1; Length 1206;
 Best Local Similarity 39.1%; Pred. No. 1.44e-01;
 Matches 18; Conservative 15; Mismatches 9; Indels 4; Gaps 4;

DB 716 VLALPNSGGPPP-PPPPPPGLAPP-PGGL-SFGLSSSSSOXP 758
 QY 98 IL-MAEGVPALPPEDAPNAASLAPTPVSPVLEPFNLTSBDDYA 142

RESULT 7 STANDARD; PRT; 1468 AA.
 AC 005860;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE FORMIN (LIMB DEFORMITY PROTEIN).
 GN FMN OR LD.
 OS MUS MUSCULUS (MOUSE).
 KW EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 -- EUTHERIA; RODENTIA.
 -- [1]
 -- SEQUENCE FROM N.A.
 -- RP TISSUE-KIDNEY AND TESTIS;
 -- RX MEDLINE: 90363291.
 -- RA WOYCHIK R.P., MAAS R.L., ZELLER R., VOGT T.F., LEDER P.;
 -- RL NATURE 346:850-853(1990).
 -- CC -1- FUNCTION: IS IMPORTANT FOR THE MORPHOGENESIS OF LIMB AND KIDNEY
 -- AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR MAY BE
 -- INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.
 -- CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 -- CC -1- TISSUE SPECIFICITY: IT IS PRESENT IN THE ADULT KIDNEY, TESTIS,
 -- LIMB, OVARY, BRAIN, SMALL INTESTINE, SALIVARY GLAND AND HARDERIAN
 -- GLAND. IT IS PRESENT THROUGHOUT THE EMBRYO.
 -- CC -1- DEVELOPMENTAL STAGE: IN THE DEVELOPING LIMB BUD, THE MESENCHYMAL
 -- IS EXPRESSED IN THE APICAL ECTODERMAL RIDGE AND THE MESENCHYMAL
 -- COMPARTMENT, PREDOMINANTLY IN THE POSTERIOR REGION. DURING
 -- THE EMBRYONAL PERIOD, EXPRESSION IS INITIALLY RESTRICTED TO
 -- THE EPITHELIAL COMPARTMENT OF THE PRONEPHROS AND MESONEPHROS.
 -- CC -1- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN
 -- ARE PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A
 -- VARIATION IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND
 -- DIFFERENT SIZE TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE.
 -- CC EMBL: X53599; G52878; -
 -- DR PIR: S11515; S11515.
 -- DR HSSP: P19999; ICLG.
 -- DR MGD: MGT:101815; FMN.
 -- KW NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING.
 -- FT DOMAIN 298 301 POLY-SER.
 -- FT DOMAIN 861 864 POLY-SER.
 -- FT DOMAIN 869 970 PRO-RICH.
 -- FT DOMAIN 977 981 POLY-SER.
 -- SQ SEQUENCE 1468 AA; 163809 MW; 42CA4104 CRC32;

Query Match 9.2%; Score 103; DB 1; Length 1468;
 Best Local Similarity 39.1%; Pred. No. 1.44e-01;
 Matches 18; Conservative 15; Mismatches 9; Indels 4; Gaps 4;

DB 942 VLALPNSGGPPP-PPPPPPGLAPP-PGGL-SFGLSSSSSOXP 984
 QY 98 IL-MAEGVPALPPEDAPNAASLAPTPVSPVLEPFNLTSBDDYA 142
 RESULT 8 STANDARD; PRT; 268 AA.
 AC 003484;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE CCAT/ENHANCER BINDING PROTEIN DELTA (C/EBP DELTA) (TRANSCRIPTION
 FACTOR CELF).

[illegible]

 W O R K I N G
 (TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1997 University of Edinburgh, U.K.
 Distribution Rights by Oxford Molecular Ltd

Mparch_un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Apr 23 05:08:39 1998; Maspar time 999.58 Seconds

Tabular output not generated. 1452.649 Million cell updates/sec

Title: >US-08-799-910-9
 Description: (1-1228) from US08799910.seq
 Perfect Score: 1228
 N.A. Sequence: 1 ATGTGCTACTCTCGCAGCTG.....AAAAAAAAAAAACTCGAG 1228
 Comp: TACACAGTAGAGACGCTGAC.....TTTTTTTTTTTGGCTC

Scoring table: TABLE default
 Gap 6

Match STD: Dbase 0; Query 0

Searched: 354530 seqs, 591221332 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database:

emb153
 1:em_in 2:em_or 3:em_om 4:em_ov 5:em_pl 6:em_htg
 7:em_hun1 8:em_hun2 9:em_ba 10:em_ro 11:em_un 12:em_v1
 13:em_pat
 genbank105
 14:gb_ro 15:gb_om 16:gb_ov 17:gb_in 18:gb_pl 19:gb_ba
 20:gb_st 21:gb_v1 22:gb_ph 23:gb_sy 24:gb_un 25:gb_pat
 26:gb_htg

Statistics: Mean 11.417; Variance 7.667; scale 1.489

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	118	9.6	1938	14	MMGLY96	1.49e+48
2	114	9.3	1758	14	RNPRL1	3.33e+46
3	79	6.4	7218	25	Sequence 14 from patent	4.27e+26
4	67	5.5	7218	25	Sequence 14 from patent	1.82e+19
5	41	3.0	215	25	Sequence 5 from patent	5.19e+06
6	37	3.0	10772	17	Drosophila melanogaster	4.03e-04
7	35	2.9	10772	17	Drosophila melanogaster	3.32e-03
8	33	2.7	128278		Sequence 5 from patent	2.61e+02
9	29	2.4	800	16	Gallus gallus fibroblast	1.36e+00
10	29	2.4	1125	15	S.scrofa mRNA for L-36	1.36e+00
11	29	2.4	1266	15	BOVIOBP	1.36e+00
12	29	2.4	1294	14	Dirofilaria immitis 70	1.36e+00
13	29	2.4	1479	14	Mus musculus mRNA for	1.36e+00
14	29	2.4	1811	16	Xenopus laevis nuclear	1.36e+00

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
15	29	2.4	1840	17	DDU73686	1.36e+00
16	29	2.4	2116	16	RCALN	1.36e+00
17	29	2.4	3042	15	SS11545	1.36e+00
18	29	2.4	3217	18	ZMR85494	1.36e+00
19	29	2.4	3737	14	S78355	1.36e+00
20	29	2.4	3789	16	XELMYBRP1	5.19e+01
21	28	2.3	571	17	SCU41441	3.51e+00
22	28	2.3	1404	16	S76875	3.51e+00
23	28	2.3	1435	17	DMU75652	3.51e+00
24	28	2.3	1008	14	MNRNASEB4	3.51e+00
25	28	2.3	2031	14	MPFRHOM	3.51e+00
26	28	2.3	2179	25	O08204	3.51e+00
27	28	2.3	2301	14	RNU75903	3.51e+00
28	28	2.3	2624	17	AF020409	3.51e+00
29	28	2.3	2628	25	166342	3.51e+00
30	28	2.3	5061	17	DDU25144	3.51e+00
31	28	2.3	3890	17	CEUC14C11	3.51e+00
32	28	2.3	58527	18	AB007649	3.51e+00
33	28	2.3	289893	26	CEY47810	3.51e+00
34	27	2.2	479	14	BMORC5	8.87e+00
35	27	2.2	714	18	BVNETVIB	8.87e+00
36	27	2.2	861	14	AF025506	8.87e+00
37	27	2.2	979	16	S7878954	8.87e+00
38	27	2.2	1444	17	HELMAMIDE	8.87e+00
39	27	2.2	2240	17	LCU56636	8.87e+00
40	27	2.2	2498	14	MAU43333	8.87e+00
41	27	2.2	2842	25	A26375	8.87e+00
42	27	2.2	3016	17	DDU31631	8.87e+00
43	27	2.2	6258	17	PFSC03080	8.87e+00
44	27	2.2	8595	17	DDU14576	8.87e+00
45	27	2.2	27471	17	CEUC44C3	8.87e+00

ALIGNMENTS

RESULT 1
 LOCUS MMGLY96 1938 bp RNA
 DEFINITION M.musculus gly96 mRNA.
 ACCESSION X67644
 NID 9287803
 KEYWORDS gly96 gene; glycosylated protein.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae; Murinae; Mus.
 1 (bases 1 to 1938)

REFERENCE
 AUTHORS Lau, L.
 TITLE Direct Submission
 JOURNAL Submitted (29-JUL-1992) L. Lau, University of Illinois College of Med., Dept. of Genetics, M/C 669, 808 South Wood Street, Chicago, IL 60612, USA
 2 (bases 1 to 1938)
 REFERENCE
 AUTHORS Charles, C.H., Yoon, J.K., Simske, J.S. and Lau, L.F.
 TITLE Genetic structure, cDNA sequence, and expression of gly96, a growth factor-inducible immediate-early gene encoding a short-lived glycosylated protein
 JOURNAL Oncogene 8 (3), 797-801 (1993)
 MEDLINE 93173526

FEATURES
 source Location/Qualifiers
 1..1938
 /organism="Mus musculus"
 /strain="BALB/c"
 /db_xref="taxon:10090"
 BASE COUNT 390 a 525 c 503 g 520 t

Query Match Best Local Similarity 9.6%; Score 118; DB 14; Length 1938;
 Matches 165; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 DB 712 ATGTGCTACTCGCAGCATCTCCACCATGCTGAGGCCCTCTTCACAGT 771
 ||||| ||||| ||||| | || ||||| ||||| ||||| |||||

QY 1 ATGTTCTACTCTGCGACGCTGCCACCCGACACATGACCATCTCTCAGACCCCGACCCCGGCG 60
DB 772 CCTCCACCGCGCGGAACTCCGGGGGCTGTGCTCCGAAATTTGACCTTCGACCT 831
QY 61 CCTCCACCATCTCCGAGACCCCGGGGGCTCGGTCTCTGAGATCTTCACCTTCGACCT 120
DB 832 CTCCCGGAGCGGCGGTGTCTCCACCCGCGCTTGAACACTTCTCCGGGCGACCGAA 891
QY 121 CTCCCGGAGCGGCGGCTCCGCGCTCCGCGCGCGCGCGCGCTCTCCGGGCGACCGAA 180
DB 892 GCGACCGAAGGCTCTCTACCTCGAGTGT 923
QY 181 GCGACCGCGAGGCTCTCTACCTCGAGTGT 212

RESULT 2
LOCUS RNRPG1 1758 bp DNA ROD 03-SEP-1996
DEFINITION R. norvegicus PRG1 gene.
ACCESSION X96437
NID 9151318
KEYWORDS PRG1 gene.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1758)
AUTHORS Schafer, H., Trauzold, A., Siegel, E.G., Polisch, U.R. and Schmidt, W.E.
TITLE PRG1: a novel early-response gene transcriptionally induced by
pituitary adenylate cyclase activating polypeptide in a pancreatic
carcinoma cell line
JOURNAL Cancer Res. 56 (11), 2641-2648 (1996)
MEDLINE 96221139
REFERENCE 2 (bases 1 to 1758)
AUTHORS Trauzold, A.
TITLE Direct Submission
SUBMITTED (05-MAR-1996) A. Trauzold, Laboratory of Molecular
Gastroenterology, 1st Dept. of Medicine, University of Kiel,
Schittenhelmstrasse 12, Kiel, D-24105, FRG
FEATURES
source
1. 1758
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/cell_line="AR4-2J"
mRNA
590..1279
/gene="PRG1"
590..1279
gene
/gene="PRG1"
BASE COUNT 358 a 475 c 470 g 455 t
ORIGIN

Query Match 9.3%; Score 114; DB 14; Length 1758;
Best Local Similarity 76.9%; Pred. No. 3.33e-46;
Matches 165; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

DB 590 ATGTGCACTTCGCTGACCACTTCGACCATGACTGCTGAGGCTCCGGCGCACTT 649
QY 1 ATGTGCACTTCGCTGACCACTTCGACCATGACTGCTGAGGCTCCGACCGCGCGCC 60
DB 650 CCTCCACCGCGCGGAACTCCGGGGGCTGTGCTCCGAAATTTGACCTTCGACCT 709
QY 61 CCTCCACCATCTCCGAGACCCCGGGGGCTCGGTCTCTGAGATCTTCACCTTCGACCT 120
DB 710 CTCCCGGAGCGGCGGTGTCTCCACCCGCGCTTGAACACTTCTCCGGGCGACCGAA 769
QY 121 CTCCCGGAGCGGCGGCTCCGCGCTCCGCGCGCGCGCGCTCTCCGGGCGACCGAA 180
DB 770 GCGACCGAAGGCTCTCTACCTCGAGTGT 801
QY 181 GCGACCGCGAGGCTCTCTACCTCGAGTGT 212

RESULT 3

LOCUS 166494 7218 bp DNA PAT 23-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
NID 92724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
source
1. 7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 5.5%; Score 67; DB 25; Length 7218;
Best Local Similarity 2.4%; Pred. No. 1.82e-19;
Matches 9; Conservative 214; Mismatches 156; Indels 0; Gaps 0;

DB 1061 TGCATATT 1120
QY 99 TGAGATCTTCACCTTCGACCTCCGCGAGCCCGCGAGCGCCCTGCGCGCGCCAG 158
DB 1121 YY 1180
QY 159 GCGCTCCGCGGCGACCGAAGCGACCGACCGAGGTTCTTACCTCGAGTGTCCGGCG 218
DB 1181 YY 1240
QY 219 CCAGCTCCGAGTCCGAGACCGAACCAGCAAGAGCTCTCTGTCTGTCTGCTACCAT 278
DB 1241 YY 1300
QY 279 CGTCTTCTGCGAGTCTCGATGAGTGAAGAGGCTGTGCGCGCCCTGCTCGACAGGA 338
DB 1301 YY 1360
QY 339 GCGCCCTAAGCGCGACATCCCTGCGCGCCGACCCCTGTGTCTCTGAGCCCTTAA 398
DB 1361 YY 1420
QY 399 TCTGACTCGAGCGCTCGAGACTAGCTGTGACCTGAGACTTCTCTCAGCAACACC 458
DB 1421 YYYYYYYYYY 1431
QY 459 GCGCGCTTCT 469

RESULT 4
LOCUS 166494 7218 bp DNA PAT 23-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
NID 92724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
source
1. 7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

DEFINITION	cds, and phenylalanyl tRNA synthetase gene, partial cds.
ACCESSION	AF012089
NID	92305220
KEYWORDS	. fruit fly.
SOURCE	Drosophila melanogaster
ORGANISM	Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 4546 to 4553)
AUTHORS	Gray, Y.H., Tanaka, M.M. and Sved, J.A.
TITLE	P-element-induced recombination in Drosophila melanogaster: hybrid element insertion
JOURNAL	Genetics 144 (4), 1601-1610 (1996)
MEDLINE	97132596
REFERENCE	2 (bases 1 to 10772)
AUTHORS	Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
TITLE	Structure of the cysteine proteinase (Cp1) gene of Drosophila melanogaster and associated mutational effects
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 10772)
AUTHORS	Gray, Y.H.M., Sved, J.A., Preston, C.R. and Engels, W.R.
TITLE	Direct Submission
JOURNAL	Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia
FEATURES	Location/Qualifiers
source	1..10772
	/organism="Drosophila melanogaster"
mRNA	/db_xref="taxon:7227"
	join(872..1000,2310..2426,6476..6690,6751..7707)
	/gene="Cp1"
	/product="cysteine protease"
gene	872..7707
	/gene="Cp1"
exon	872..1000
	/gene="Cp1"
	/number=1
intron	1001..2309
	/gene="Cp1"
exon	2310..2426
	/gene="Cp1"
	/number=2
cds	join(2328..2426,6476..6690,6751..7462)
	join(2328..2426,6476..6690,6751..7462)
	/gene="Cp1"
	/codon_start=1
	/product="cysteine proteinase-1"
	/db_xref="PID:g2305321"
	/translation="MRTAVLPLALIAVQAQSFADVYMEEKHFKLEHRKNYDEEERFLKLFENKRIKAKHNOFAEGKVSRLVANKYALDLHHEFROLNNGENVYLLHQLRADESKGVGTIFISPAHYTLPRSYDMRTGAVTAVKDOGHGSCMAISGALGEALHRSKGVIVLSLEONLVDGSKYKGNNGNGLMDNMFYIKDNGIDTEKSYPEYAIIDHCKRGTGATDRGFTDIPQGEKKAQAVATVGPVSVAIDASHSEPOFSEGVYDPOCAQNLHDHGVLYVGFDTDESEDIWLYKNSGTTWGDKGIKLRKKEQCIQIAAASSIPLYV
intron	2427..6475
	/gene="Cp1"
misc_feature	4546..4553
	/citation=[1]
	6476..6690
exon	/gene="Cp1"
	/number=3
intron	6691..6750
	/gene="Cp1"
exon	6751..7707
	/gene="Cp1"
	/number=4
mRNA	join(8110..9300,9370..>9532)
	/product="phenylalanyl tRNA synthetase"
cds	join(8110..9300,9370..>9532)
	/note="Potential orf"

FEATURES	source
gene	872..7707 /product="cysteine protease"
exon	872..1000 /gene="Cpi"
intron	1001..2309 /gene="Cpi"
exon	2310..2426 /gene="Cpi"
exon	2427..2596 /number=2
exon	2597..2707 /number=1
exon	2708..2818 /number=1
exon	2819..2929 /number=1
exon	2930..3040 /number=1
exon	3041..3151 /number=1
exon	3152..3262 /number=1
exon	3263..3373 /number=1
exon	3374..3484 /number=1
exon	3485..3595 /number=1
exon	3596..3706 /number=1
exon	3707..3817 /number=1
exon	3818..3928 /number=1
exon	3929..4039 /number=1
exon	4040..4150 /number=1
exon	4151..4261 /number=1
exon	4262..4372 /number=1
exon	4373..4483 /number=1
exon	4484..4594 /number=1
exon	4595..4705 /number=1
exon	4706..4816 /number=1
exon	4817..4927 /number=1
exon	4928..5038 /number=1
exon	5039..5149 /number=1
exon	5150..5260 /number=1
exon	5261..5371 /number=1
exon	5372..5482 /number=1
exon	5483..5593 /number=1
exon	5594..5704 /number=1
exon	5705..5815 /number=1
exon	5816..5926 /number=1
exon	5927..6037 /number=1
exon	6038..6148 /number=1
exon	6149..6259 /number=1
exon	6260..6370 /number=1
exon	6371..6481 /number=1
exon	6482..6592 /number=1
exon	6593..6703 /number=1
exon	6704..6814 /number=1
exon	6815..6925 /number=1
exon	6926..7036 /number=1
exon	7037..7147 /number=1
exon	7148..7258 /number=1
exon	7259..7369 /number=1
exon	7370..7480 /number=1
exon	7481..7591 /number=1
exon	7592..7702 /number=1
exon	7703..7813 /number=1
exon	7814..7924 /number=1
exon	7925..8035 /number=1
exon	8036..8146 /number=1
exon	8147..8257 /number=1
exon	8258..8368 /number=1
exon	8369..8479 /number=1
exon	8480..8590 /number=1
exon	8591..8701 /number=1
exon	8702..8812 /number=1
exon	8813..8923 /number=1
exon	8924..9034 /number=1
exon	9035..9145 /number=1
exon	9146..9256 /number=1
exon	9257..9367 /number=1
exon	9368..9478 /number=1
exon	9479..9589 /number=1
exon	9590..9700 /number=1
exon	9701..9811 /number=1
exon	9812..9922 /number=1
exon	9923..10033 /number=1
exon	10034..10144 /number=1
exon	10145..10255 /number=1
exon	10256..10366 /number=1
exon	10367..10477 /number=1
exon	10478..10588 /number=1
exon	10589..10699 /number=1
exon	10700..10810 /number=1
exon	10811..10921 /number=1
exon	10922..11032 /number=1
exon	11033..11143 /number=1
exon	11144..11254 /number=1
exon	11255..11365 /number=1
exon	11366..11476 /number=1
exon	11477..11587 /number=1
exon	11588..11698 /number=1
exon	11699..11809 /number=1
exon	11810..11920 /number=1
exon	11921..12031 /number=1
exon	12032..12142 /number=1
exon	12143..12253 /number=1
exon	12254..12364 /number=1
exon	12365..12475 /number=1
exon	12476..12586 /number=1
exon	12587..12697 /number=1
exon	12698..12808 /number=1
exon	12809..12919 /number=1
exon	12920..13030 /number=1
exon	13031..13141 /number=1
exon	13142..13252 /number=1
exon	13253..13363 /number=1
exon	13364..13474 /number=1
exon	13475..13585 /number=1
exon	13586..13696 /number=1
exon	13697..13807 /number=1
exon	13808..13918 /number=1
exon	13919..14029 /number=1
exon	14030..14140 /number=1
exon	14141..14251 /number=1
exon	14252..14362 /number=1
exon	14363..14473 /number=1
exon	14474..14584 /number=1
exon	14585..14695 /number=1
exon	14696..14806 /number=1
exon	14807..14917 /number=1
exon	14918..15028 /number=1
exon	15029..15139 /number=1
exon	15140..15250 /number=1
exon	15251..15361 /number=1

[illegible]

	9		VRT.	02-APR-1996
LOCUS	GGU41467	800 bp	mRNA	
DEFINITION	Gallus gallus fibroblast growth factor 8 FGF8 mRNA, partial cds.			
ACCESSION	U91467			
NID	g1134969			
KEYWORDS	.			
SOURCE	chicken.			
ORGANISM	Gallus gallus Eukaryote; mitochondria eukaryotes; Metazoa; Chordata; Vertebrata; Archosauria; Aves; Neognathae; Galliformes;			
	Phasianidae; Phasianinae; Gallus.			
REFERENCE	1 (bases 1 to 800)			
AUTHORS	Crossley,P.H., Minowada,G., MacArthur,C.A. and Martin,G.R. Roles for FGFR in the induction, initiation, and maintenance of chick limb development <i>Cell</i> 84 (1), 127-136 (1996)			
TITLE	JOURNAL MEDLINE	96140646		
	2 (bases 1 to 800)			
REFERENCE	Crossley,P.H., Minowada,G., MacArthur,C. and Martin,G.R. Direct Submission Submitted (27-NOV-1995) Philip H. Crossley, Anatomy, UCSF, Parnassus, San Francisco, CA 94143-045, USA			
AUTHORS	Title	Location/Qualifiers		
JOURNAL		I . 800		
FEATURES				
Source				

	RESULT	10			
LOCUS	SSI36ABP	1125 bp	RNA	MAM	11-JAN-1995
DEFINITION	S.scrofa mRNA for L-36 lactose binding lectin.				
ACCESSION	X79303				
NID	9623345				
KEYWORDS	1-361bp gene; lactose-binding lectin.				
SOURCE	pig				
ORGANISM	Sus scrofa				
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Artiodactyla; Suidae; Sus.				
AUTHORS	1 (bases 1 to 1125)				
TITLE	Chiu,M.L., Perry,D.A., Feldman,S.R., Klapper,D.G. and O'Keefe,E.J. An adherens junction protein is a member of the family of lactose-binding lectins				
JOURNAL	J. Biol. Chem.	269 (50),	31770-31776	(1994)	
MEDLINE	95081129				
REFERENCE	2 (bases 1 to 1125)				
AUTHORS	O'Keefe,E.J.				
TITLE	Direct Submission				
JOURNAL	Submitted (18-MAY-1994) E.J. O'Keefe, University of North Carolina 137 NCHM, Dept of Dermatology, CB7600, Chapel Hill, NC 27514, USA				
FEATURES	location/Qualifiers				
source	1..1125				

Query Match 2.4%; Score 29; DB 15; Length 1125;
Best Local Similarity 89.2%; Pred. No. 1,35e+00;
Matches 33; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 1089 AATAAAATTATAATAGCCAAAAAAAAAAAAAAAAAAAA 1125
||||| | | | | | | | | | | | | | | | | | | | |
QY 1185 AATAAAATTGATTACTACTGTCAAAAAAAAAAAAAAAAAAAAA 1221

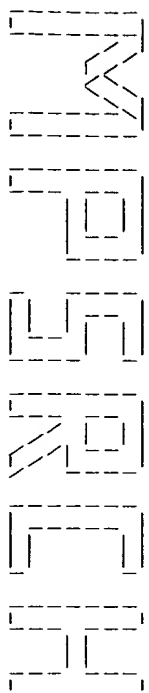
RESULT 11
LOCUS BOV10PPP 1266 bp mRNA MAM 25-JAN-1993
DEFINITION Bovine inorganic pyrophosphatase mRNA sequence.
ACCESSION M95283
NID g163228
KEYWORDS inorganic pyrophosphatase.
SOURCE Bos taurus retina cDNA to mRNA.
ORGANISM Bos taurus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 1266)
Yang, Z. and Wensel, T. G.
Molecular cloning and functional expression of cDNA encoding a
mammalian inorganic pyrophosphatase
J Biol. Chem. 267, 24641-24647 (1992)
MEDLINE JOURNALL 93077539

FEATURES		Location/Qualifiers
SOURCE		1..1266 /organism="Bos taurus" /db_xref="taxon:9913" /tissue_type="retina"
BASE COUNT	396 a 239 c 300 g 331 t	
ORIGIN		
Query Match	2.4%; Score 29; DB 15; Length 1266;	
Best Local Similarity	87.2%; Pred. No. 1.36e+00;	
Matches	34; Conservative 0; Mismatches 5; Indels 0; Gaps 0;	
Db	1227 AATATAAATTATTTGTTCAAAAAAAAAAAAAAA 1265	
Oy	1184 AATATAAATTGATTTACTGTCTCAAAAAAAAAAAAA 1222	
RESULT 12		
LOCUS	DIRHSP70 1294 bp mRNA INV 25-AUG-1992	
DEFINITION	Dirofilaria immitis 70 kda heat shock protein homologue (hsp 70)	
ACCESSION	M95648	
NID	g156705	
KEYWORDS	heat shock protein 70; heat shock protein homologue.	
SOURCE	Dirofilaria immitis adult cDNA to mRNA.	
ORGANISM	Eukaryotes; mitochondrial eukaryotes; Metazoa; Nematoda; Secernentea; Spirurida; Sphuridia; Spirulina; Filarioidea; Onchocercidae; Dirofilaria.	
REFERENCE	1 (bases 1 to 1294) Culpepper,J.A., Friedman,L. and Dale,B. Molecular cloning and characterization of a Dirofilaria immitis cDNA encoding an Hsp 70 homologue Unpublished (1992)	
JOURNAL	location/qualifiers	
FEATURES		
source	1..1294 /organism="Dirofilaria immitis" /db_xref="taxon:6287" /dev_stage="adult" 1..1038 /partial /note="homologue" /codon_start=1 /product="heat shock protein 70" /db_xref="pid:g156706" /translation:"RPPELCDDFRSTMDPYEKALRDANKRAVHDIVLYGGSTRIFKVGKLSDPFGSGELKNSINDDEVAITGAAYQAILSGDSSEAVODLLDLVAFLSLGLETRGGVWTAALKRNTTIPITSQTFTYSONOGVLIQVEGERAMTKDNLLGKEE LSGTPAPRGVAPOLIEVTIDANGILNWSADCKTGOKKITINDGRSLSKDEIERHVOEAERKKADDEAKDRIAAKNALESYAFNMKRIEDEKLDKISEDDKKIKIOKCDGE TVRWLNDNOTAEKDEFHFHROKLELSVCNPITTKIYQSAGMGPMGPMPGCARGCGSS TGGGPTLEEVD"	
CDS		
BASE COUNT	409 a 223 c 296 g 360 t	
ORIGIN		
Query Match	2.4%; Score 29; DB 17; Length 1294;	
Best Local Similarity	87.2%; Pred. No. 1.36e+00;	
Matches	34; Conservative 0; Mismatches 5; Indels 0; Gaps 0;	
Db	1256 AAGTAAATTTATATGCGGTCAAAAAAAAAAAAAA 1294	
Oy	1184 AATATAAATTGATTTACTGTCTCAAAAAAAAAAAAA 1222	
RESULT 13		
LOCUS	MM26SPROT 1479 bp RNA ROD 07-OCT-1997	
DEFINITION	Mus musculus mRNA for 26S proteasome non-ATPase subunit.	
ACCESSION	y13071	
NID	q2505939	
KEYWORDS	26S proteasome; non-ATPase subunit; proteolysis; regulatory complex.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
gene
CDS
BASE COUNT
ORIGIN
Query Match
Best Local Similarity 87.2%;
Matches 34; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db 1440 AAATAAATACATTACATCCAAAAA 1478
Ox 1184 AAATAAATGATTACTCAAAAAA 1222
RESULT 14
LOCUS
DEFINITION
ACCESSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. 1479
/organism="Mus musculus"
/db_xref="taxon:10090"
/dev_stage="adult"
237..1166
/gene="PADI"
237..1166
/gene="PADI"
/function="proteolysis"
/codon_start=1
/product="26S proteasome, non-ATPase subunit"
/db_xref="PID:e352083"
/db_xref="PID:g2505940"
/translation="MDRLRLRGSGMGLARPPTDPAVDPAEYVYSSILLLKMLKKG
RAGYPMVMGLMGFEYDDTVRYIDYDFAPDQSTGYSEVADYDFPAKMLDKQKG
RPEAVYGMHSHPEFGCWLGSVDINTQOSEALSERAAVAVDQVSKGVADAR
LINAAMVLGHEPRTTSLNGLKRPISQALHGLNRYYSITINPKNEQKMLLN
LHKSMEEGLLDQYSECHKNESVSEMELELANYNKAAVEEDKMPEDLAINKVK
ODPRLHEEHVDVMTSNIVQCLAAMLDTVFK"
African clawed frog.
Xenopus laevis
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea;
Pipidae; Xenopodine; Xenopus.
1 (bases 1 to 1811)
Cordes,V.C., Reidenbach,S., Rackwitz,H.R. and Franke,W.W.
Identification of protein p270/Tpr as a constitutive component of
the nuclear pore complex-attached intranuclear filaments
J. Cell Biol. 136 (3), 515-529 (1997)
2 (bases 1 to 1811)
Cordes,V.C., Hunzicker,A. and Franke,W.W.
Direct Submission
Submitted (06-SEP-1996) Cell Biology/0110, German Cancer Research
Center, INF 280, Heidelberg 69120, Germany
Location/Qualifiers
1..1811
/organism="Xenopus laevis"
/db_xref="taxon:8355"

```

(TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

Msearch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Apr 23 05:39:41 1998; Maspar time 157.52 Seconds

Tabular output not generated. 899.629 Million cell updates/sec

Title: >US-08-799-910-9

Description: (1-1228) from US08799910.seq

Perfect Score: 1228

N.A. Sequence: 1

Comp: 1 ATGCTGCTACTCTCCGACGCTG.....AAAAAATGCTGCTG 1228
TACACAGTGAAGAGCGTGCAC.....TTTTTTTTTTTGTGACCTC

Scoring table: TABLE default

Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 159651 seqs, 57698962 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

n.geneseq30
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33

Statistics: Mean 9.360; Variance 6.778; scale 1.381

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	135	11.0	297	10	Human genome fragment	1.11e-57
2	94	7.7	1047	2	Human Natriuretic Pep	5.93e-35
3	80	6.5	1047	2	Human Natriuretic Pep	2.20e-27
4	62	5.0	267	10	Human genome fragment	6.80e-18
5	47	3.8	53	18	Human gene signature	2.61e-10
6	46	3.7	91	9	Oligonucleotide probe	8.07e-10
7	43	3.5	91	9	Oligonucleotide probe	2.31e-08
8	43	3.5	204	1	Base substituted E.co	2.10e-07
9	41	3.3	204	1	Base substituted E.co	1.59e-05
10	37	3.0	114	12	Generic DNA sequence	1.32e-04
11	35	2.9	114	12	Generic DNA sequence	4.59e-05
12	36	2.9	114	12	Generic DNA sequence	4.59e-05
13	36	2.9	114	12	Generic DNA sequence	4.59e-05
14	36	2.9	114	12	Generic DNA sequence	4.59e-05
15	35	2.9	178	31	Human endothelin-1 an	1.32e-04

C	16	34	2.8	100	31	T76186	Human IL4 receptor an	3.74e-04
C	17	34	2.8	114	12	Q70465	Generic DNA sequence	3.74e-04
C	18	34	2.8	114	12	Q70466	Generic DNA sequence	3.74e-04
C	19	34	2.8	114	12	Q70470	Generic DNA sequence	3.74e-04
C	20	34	2.8	114	12	Q70465	Generic DNA sequence	3.74e-04
C	21	34	2.8	114	12	T76233	Human IL6 antisense o	1.05e-03
C	22	33	2.7	114	12	Q70469	Generic DNA sequence	1.05e-03
C	23	33	2.7	178	31	T76405	Human endothelin-1 an	2.93e-03
C	24	33	2.6	114	12	Q70467	Generic DNA sequence	2.93e-03
C	25	32	2.6	114	12	Q70468	Generic DNA sequence	2.93e-03
C	26	32	2.6	172	32	T76363	Human Interleukin 8 a	2.93e-03
C	27	31	2.5	114	12	Q70470	Generic DNA sequence	8.08e-03
C	28	31	2.5	2420	7	Q46672	p mySA cDNA clone.	5.92e-02
C	29	29	2.4	88	31	T76170	Human IL3 receptor an	2.20e-02
C	30	30	2.4	100	31	T76186	Human IL4 receptor an	2.20e-02
C	31	30	2.4	114	12	Q70473	Generic DNA sequence	2.20e-02
C	32	30	2.4	114	12	Q70471	Generic DNA sequence	2.20e-02
C	33	29	2.4	114	12	Q70472	Generic DNA sequence	2.20e-02
C	34	29	2.4	190	31	T76452	Chymase antisense o1	5.92e-02
C	35	29	2.4	250	31	T76438	Substance P antisense	1.57e-01
C	36	28	2.3	114	12	Q70471	Generic DNA sequence	1.57e-01
C	37	28	2.3	130	31	T76152	Human vascular cell a	1.57e-01
C	38	28	2.3	162	31	T76307	Human RANTRIS antisens	1.57e-01
C	39	28	2.3	200	31	T76398	Human Leukotriene C4	1.57e-01
C	40	28	2.3	595	19	T16988	E-Dex integrin inhibi	1.57e-01
C	41	28	2.3	2179	13	Q73117	Alpha 2, 3-sialyl tra	1.57e-01
C	42	28	2.3	2232	12	Q77739	Human alpha-2,3-sialy	1.57e-01
C	43	28	2.3	2292	13	Q79934	Human liver hap cDNA.	1.57e-01
C	44	28	2.3	2992	20	T06491	Human hepatoma celli	1.57e-01
C	45	28	2.3	2992	11	Q65572	Human liver hap cDNA.	1.57e-01

ALIGNMENTS

RESULT	1	Q77534 standard; DNA: 297 BP.
ID	Q77534;	AC 23-SEP-1994 (first entry)
DE	Human genome fragment. (Preferred)	
KW	Brain; Placenta; bone marrow; genetic analysis; gene mapping;	
KW	detection; homology; human; adrenal tissue; ds.	
OS	Homo sapiens.	
PN	WO9401548-A.	
PF	20-JAN-1994.	
PR	13-JUL-1993; G01467.	
PA	13-JUL-1992; GB-014857.	
PI	Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;	
PI	Sibson DR, Starkey M;	
DR	MP1: 94-035056/04.	
PT	New nucleic acid fragment encoding gene products - can be used	
PT	for genetic analysis and mapping	
PS	Claim 1; Page 575-576; 61bp; English.	
CC	Human nucleic acid fragments, isolated from brain, adrenal tissue,	
CC	the placenta or bone marrow comprise any of: (A) a sequence	
CC	selected from (Q76401-Q77613), (B) an allelic variation of a	
CC	sequence as described in (A), or (C) a sequence complementary	
CC	to (A) or (B).	
CC	Preferred sequences exhibit no more than 90105mology to a human	
CC	sequence known per se.	
CC	Sequence 297 BP; 66 A; 117 C; 61 G; 52 T;	
CC	Sequence	
Query Match	11.0%; Score 135; DB 10; Length 297;	
Best Local Similarity	76.6%; Pred. No. 1.11e-57;	
Matches	209; Conservative 0; Mismatches 62; Indels 2; Gaps 2;	
Db	22 cactcgcgcgtgcacaccccaatgacacccctacagaccccaacccacccctac 81	
QY		
QY	7 cactctgcgacgtccacccgaacacacacacacacacacacacacacacacac 66	
Db	82 gtatctgtccac 140	
QY		
QY	67 ACCATCCCGGACCCCGGCGGGGCTCCGCTCCTGAGATCTTACCTTCACCTCTCCG 126	

[illegible]

```

Cc      700 CTCCTTCCACAGGGGCGCTAGCCCGACAGCTGGGGCTGGCTGGCTCTATAGGGCTCGTC
Dd      268 ratg-rmwndrtrnnanannannatvrvltynnnnnnynnnnnnnnrnnarndngvng
Cc      640 TCTGGCCCTCGGCTCCCGCTCAAGCACGGGGGGCTGCTCCGCTAGTGTCTAGT
Dd      327 nsmnnnagcnydygnnyanvrnnrltnnggrindgrnrvn-kmgrryhytgvrwvdkn
Cc      560 TCAAGTTCCTCGGAAGTCCACGATTGGGGATACGCTCTCGGCCCGCAGTACGCTGTG
Dd      386 ndrntdvnmagmdsdsgdnnaahysganrnmwvtgrnnrvngyansdnndnandd
Cc      520 TTTCTTTGTGGTTTTTCGGATTCTTTTGGGGAATCGCGGAGTCAAGTTAAGAACGGC
Dd      446 nscdktnstnanvanygtntnrmvysnnnrkmmnknnnasnmrvrvnnnnngsn
Cc      460 CCGGGATTGCTGGAGGAATGCTGAGCTCAAGCTAGTATCCGAGGGCTCGAATCA
Dd      506 rthkxa-gsrtnsnrg-ssygsmtahkyrnnantghknrvanrhvnhktrnnltin
Cc      400 GATTAAAGGCTCGAGCAGGGGACACAGGGGTGGCGCCAGGAGATGGCGCTTAGGGC
Dd      564 vnnnnkhmrdrnnnhtnngnacdnnnnnrvtnyrcrnsndnnndnnndmrvysn
Cc      340 CGTCTCTGGAGAGCAGGGGGCCCCCGCACACCTCTTCAAGCATAGATCGCAGAGA
Dd      624 ndnrvkgnanhnshnsghsnksncvdsrtnvknktdyngnaestrlanddnnnyak
Cc      280 CGATGTGATGACAGA-GAAGAAGAACCTTTGGCTGGGTTCGGTTC-TCGACTGGCAG
Dd      684 knntannnnsgnnnttymnaadyvsngnnnnnnnansgnyngndsnknvnhkvrng
Cc      222 CTGGCGCCGGACACATCGAGGGTA-GAAGAACCTCGCGCTCGCTTCGGGGCCCCGAG
Dd      744 nrryrrnsdtrnnnnnnvnmrcwandaanrdngnkxnrrnnkngtstndnnn
Cc      163 AGCGCTCGGGCGCCCGCAGGGCGCTCGGGCTCCGGAGAGGTCGAAGTGAAGA
Dd      804 rmmnyannnhkvnnrtlaynnhkrkananynnnhsvannhkrnltvnanandsvny
Cc      103 TCTCAGACCGAGCGCCCGCGGGTCCCGGGATGTGTGAAGGGGGCGGGGCT
Dd      902 nsdvngntansanstnmrvltnndnytcndandndv
Cc      43 GCAGATGTCTATGTCTCGGGTGCAGCTCGAGAGTGC
Cc      RESULT 4
Cc      ID 077554 standard; DNA: 267 BP.
Cc      AC 077554:
Cc      DT 23-SEP-1994 (first entry)
Cc      DE Human genome fragment. (Preferred)
Cc      KW Brain, placenta, bone marrow, genetic analysis; gene mapping;
Cc      KW detection; homology; human; adrenal tissue; ds.
Cc      OS Homo sapiens.
Cc      PN WO9401548-A.
Cc      PD 20-JAN-1994.
Cc      PF 13-JUL-1993: G01467.
Cc      PR 13-JUL-1992; GB-014857.
Cc      PA (MEDICAL RES COUNCIL.
Cc      PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D:
Cc      PI Gibson DR, Starkey M;
Cc      PI WPI: 94-035056/04.
Cc      PT New nucleic acid fragment encoding gene products - can be used
Cc      PT for genetic analysis and mapping
Cc      PS Claim 1; Page 586; 616pp; English.
Cc      CC Human nucleic acid fragments, isolated from brain, adrenal tissue,
Cc      CC the placenta or bone marrow, comprise any of: (A) a sequence
Cc      CC selected from (076401-077613), (B) an allelic variation of a
Cc      CC sequence as described in (A), or (C) a sequence complementary
Cc      CC to (A) or (B).
Cc      CC Preferred sequences exhibit no more than 90105mology to a human
Cc      CC sequence known per se.

```

```

SQ Sequence 267 BP; 58 A; 117 C; 49 G; 42 T;
Query Match 5.08; Score 62; DB 10; Length 267;
Best Local Similarity 75.08; Pred. No. 6.80e-18;
Matches 114; Conservative 0; Mismatches 35; Indels 3; Gaps 3;

Db 116 ccggaatcttcaccccttagacccttcctccgaaacccgcagccgccttc-gaaccttc 174
|||||
OY 97 cctgaatcttcaccccttagacccttcctccgaaacccgcagccgccttc-gaaccttc 156
|||||

Db 175 aactcgt-tcaacagcgccgaaagcctatccagagactctcgcgcctcagtg-a-cgpn 232
|||||
OY 157 AGCGCCTCTCCGCGCAGCCCAAGCCGAGCCGAGGTTCTCTACCTCGATGTCGG 216

Db 233 ccctagctcgtcagtcgagaatccgaactccgc 264
|||||
OY 217 CGCCAGCTGCAGCTCGAGAACCGAACCCAGC 248

RESULT 5
ID T20789 standard; cDNA to mRNA; 53 BP.
AC T20789;
DT 09-JUL-1996 (first entry)
DE Human gene signature HUMGS02013.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994: J01916.
PR 12-NOV-1993: JP-355504.
PA (MATS/) MATSUBARA K.
PI (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues.
PS Claim 1: Page 723; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in 119001-126837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 53 BP; 18 A; 9 C; 9 G; 17 T;

Query Match 3.88; Score 47; DB 18; Length 53;
Best Local Similarity 100.0%; Pred. No. 2.61e-10;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gatcgctagatgtctctgtaacacaataaattgattactgtc 47
|||||
OY 1158 GATCGCTAGATGTCTCTGTGAACACAATAAATTGATTACTGTGC 1204

RESULT 6
ID 051746 standard; cDNA; 91 BP.
AC 051746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;

KM OS.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993: 108325.
PR 26-MAY-1992: US-889651.
PA (BECT) BECTON DICKINSON CO.
PI Shank DD, Spears PA.
DR WPI: 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples.
PS Claim 3: Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (051735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also 051735-45 and 051747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 3.78; Score 46; DB 9; Length 91;
Best Local Similarity 13.3%; Pred. No. 8.07e-10;
Matches 8; Conservative 45; Mismatches 7; Indels 0; Gaps 0;

Db 1 ggctccggcgssvhyvvvshvshvshvshvshvshvshvshvshvshvshvshvsv 60
|||||
OY 88 GGCTCCGGCTCGAGATCTTCACCTTCGACCTCCGAGCCGAGCCGAGCCGCTCC 147

RESULT 7
ID 051746 standard; cDNA; 91 BP.
AC 051746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW OS.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993: 108325.
PR 26-MAY-1992: US-889651.
PA (BECT) BECTON DICKINSON CO.
PI Shank DD, Spears PA.
DR WPI: 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples.
PS Claim 3: Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (051735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also 051735-45 and 051747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 3.5%; Score 43; DB 9; Length 91;
Best Local Similarity 2.0%; Pred. No. 2.31e-08;
Matches 1; Conservative 45; Mismatches 3; Indels 0; Gaps 0;

Db 13 vhsyvvvshvshvshvshvshvshvshvshvshvshvshvshvshvshvshvsv 61
|||||
Cp 1153 CAGTCGCTCGTGACAGCCACCAACCCCAACCTTCACCTCGAGCC 1105

RESULT 8
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
PH Key misc_feature 19..69

```


RESULT 13
ID 070467 standard: DNA: 114 BP.
AC 070467;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc-feature 55..60
FT /*tag= a
FT /note= "this sequence represents '2'; 2 can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PR 01-FEB-1994; US-013416.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB: R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PS Disclosure; Page 35; 255pp; English.
CC 070467 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)16(TGC)(NNB)1Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in 070466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed activity allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;
Query Match 2.9%; Score 36; DB 12; Length 114;
Best Local Similarity 3.7%; Pred. No. 4.59e-05;
Matches 4; Conservative 33; Mismatches 71; Indels 0; Gaps 0;

KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc-feature 55..60
FT /*tag= a
FT /note= "this sequence represents '2'; 2 can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PR 01-FEB-1994; US-013416.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB: R65152.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PS Disclosure; Page 35; 255pp; English.
CC 070466 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)1(TGC)(NNB)14(TGC)(NNB)8(TGC)(NNB)
CC -9Y. X and Y are flanking restriction sites (X is not the same as Y)
CC that are not specified further. Other generic sequences are shown in
CC 070466-68. Other specific peptides generated by these generic sequences
CC are shown in R65151-54. TSARs are concatenated heterofunctional proteins
CC or peptides, comprising at least two functional regions - a binding
CC domain with affinity for a ligand and a second effector peptide portion
CC that is chemically or biologically active. They may further comprise a
CC linker peptide between the 2 domains. The oligonucleotides are also
CC designed so that the expressed peptide contains 2 or 4 cysteine residues
CC positioned in, or flanking, the unpredicted or variant residues. These
CC residues confer some degree of conformational rigidity to the peptides.
CC The TSARs or compns. comprising a TSAR binding domain can be used in
CC vivo to deliver a chemically or biologically active moiety, eg. metal
CC ion, radioisotope, peptide, toxin or enzyme, to the specific target or
CC on the cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed activity allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;
Query Match 2.9%; Score 36; DB 12; Length 114;
Best Local Similarity 7.3%; Pred. No. 4.59e-05;
Matches 8; Conservative 31; Mismatches 71; Indels 0; Gaps 0;

Db 5 gcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnnnnn 64
CP 703 GCTCTCTCTCCACCGGCGCTAGCCGAGCTGCTGCTCTATGCGCCTCG 644
Db 65 nbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 114
CP 643 GTCCTGTGCGCTGCTCCGCTCAAGCAGCGGTGGCTCTCCGCTG 594

RESULT 15
ID T76405 standard: DNA: 178 BP.
AC T76405;
DT 15-SEP-1997 (first entry)
DE Human endothelin-1 antisense oligonucleotide.
KW Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW chronic obstructive pulmonary disease; bronchitis; ss.
OS Synthetic.
PN WO9640162-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; US-009306.
PR 07-JUN-1995; US-474497.

DT 06-JUL-1995 (first entry)
 DE Human NMDAR2 receptor subunit clone NMDA21 protein.
 KW N-methyl-D-aspartate; receptor; human; NMDA; cation-selective; channel;
 KM glutamate; hippocampus; rat; pcdNAL; NMDA receptor; antagonist.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc_difference 473..474
 FT /note- "deletion of 17 a.a. from NMDAR2C a.a.
 FT sequence between these residues"
 PN WO9424284-A.
 PD 27-OCT-1994.
 PF 20-APR-1994; U04387.
 PR 20-APR-1993; US-052449.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Daggett LP, Ellis SB, Liaw CM, Lu C;
 DR WPI: 94-341863/42.
 DR N-PSDB: Q79399.
 PT Isolated DNA encoding a human N-methyl-D-aspartate receptor
 PT subunit - used as probes in the identification and isolation of
 PT nucleic acids encoding related receptor subunits.
 PS Claim 11: Page 99-104; 156pp: English.
 CC The amino acid sequence of the novel N-methyl-D-aspartate (NMDA) receptor
 CC 2C (NMDAR2C) gene clone NMDA21. This clone covers the a.a. 249-1040 of
 CC the complete NMDAR2C protein (Q79372) with the exception of a 17 a.a.
 CC deletion of a.a. 722-738 of the NMDAR2C sequence. The NMDA receptor
 CC contains two subunits: subunit R1 (Q79370) and subunit R2 selected from
 CC the subunits 2A (Q79375), 2B (Q79377), 2C and 2D (Q79378). The receptor
 CC forms part of a family of NMDA receptors which have cation-selective
 CC channels and bind glutamate and NMDA. The NMDAR2C gene was obtained by
 CC amplifying cDNA derived from human brain tissues with primers
 CC corresponding to sequences in the rat NMDAR2A receptor DNA and using the
 CC resultant fragments as probes to screen a cDNA library derived from human
 CC hippocampal RNA. 4 basic clones were isolated: NMDA21 (Q79399), NMDA22
 CC (Q79400), NMDA24 (Q79401) and NMDA26 (Q79402). The clones are thought to
 CC be splice variants of each other. Based on the sequence of the 4 clones,
 CC a series of variants (Q79403-7) of the NMDAR2C receptor were constructed.
 CC The expression of the genes allows the reconstruction of the NMDA
 CC receptor. The complete receptor can be used to identify compounds which
 CC bind or are antagonistic to the human NMDA receptor.
 CC Sequence 775 AA;
 SO

Query Match 8.6%; Score 96; DB 12; Length 775;
 Best Local Similarity 42.5%; Pred. No. 1.78e+01;
 Matches 17; Conservative 11; Mismatches 12; Indels 0; Gaps 0;
 Db 664 appspcptrgspgpcldpdppepsctgwgppdgqra 703
 QY 18 TPAPSTIPGRSGPEITFDLPDEPAAPAGRPASRG 57

RESULT 9
 ID R66065 standard; Protein; 1214 AA.
 AC R66065;
 DT 06-JUL-1995 (first entry)
 DE Human NMDAR2 receptor subunit clone NMDAR2C-delta15-delta1 gene.
 KW N-methyl-D-aspartate; receptor; human; NMDA; cation-selective; channel;
 KM glutamate; hippocampus; rat; pcdNAL; NMDA receptor; antagonist.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc_difference 591..592
 FT /note- "5 a.a. deletion of NMDAR2C a.a. sequence
 FT between these residues"
 FT Misc_difference 716..717
 FT /note- "17 a.a. deletion of NMDAR2C a.a. sequence
 FT between these residues"
 PN WO9424284-A.
 PD 27-OCT-1994.
 PF 20-APR-1994; U04387.
 PR 20-APR-1993; US-052449.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Daggett LP, Ellis SB, Liaw CM, Lu C;
 DR WPI: 94-341863/42.
 DR N-PSDB: Q79407.

PT Isolated DNA encoding a human N-methyl-D-aspartate receptor
 PT subunit - used as probes in the identification and isolation of
 PT nucleic acids encoding related receptor subunits.
 PS Claim 11: Page 99-104; 156pp: English.
 CC The amino acid sequence of the novel N-methyl-D-aspartate (NMDA) receptor
 CC 2C (NMDAR2C) gene clone NMDAR2C-delta15-delta1. This clone covers the
 CC same sequence as the NMDAR2C sequence but contains deletions of a.a.
 CC 591-5 and 722-38 of the NMDAR2C protein (R66035). A series of variants
 CC (Q79403-7) of the NMDAR2C receptor were constructed. By recombination,
 CC based on the sequence of the 4 basic clones: NMDA21 (Q79399), NMDA22
 CC (Q79400), NMDA24 (Q79401) and NMDA26 (Q79402). The NMDA receptor
 CC contains two subunits: subunit R1 (Q79370) and subunit R2 selected from
 CC the subunits 2A (Q79375), 2B (Q79377), 2C and 2D (Q79378). The receptor
 CC forms part of a family of NMDA receptors which have cation-selective
 CC channels and bind glutamate and NMDA. The expression of the genes allows
 CC the reconstruction of the NMDA receptor. The complete receptor can be
 CC used to identify compounds which bind or are antagonistic to the human
 CC NMDA receptor.
 CC Sequence 1214 AA;
 SO

Query Match 8.6%; Score 96; DB 12; Length 1214;
 Best Local Similarity 42.5%; Pred. No. 1.78e+01;
 Matches 17; Conservative 11; Mismatches 12; Indels 0; Gaps 0;

Db 907 appspcptrgspgpcldpdppepsctgwgppdgqra 946
 QY 18 TPAPSTIPGRSGPEITFDLPDEPAAPAGRPASRG 57

RESULT 10
 ID R66063 standard; Protein; 1219 AA.
 AC R66063;
 DT 06-JUL-1995 (first entry)
 DE Human NMDAR2 receptor subunit clone pcMV-26-scat-21-NotI-24 protein.
 KW N-methyl-D-aspartate; receptor; human; NMDA; cation-selective; channel;
 KM glutamate; hippocampus; rat; pcdNAL; NMDA receptor; antagonist.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc_difference 721..722
 FT /note- "17 a.a. deletion of NMDAR2C a.a. sequence
 FT between these residues"
 PN WO9424284-A.
 PD 27-OCT-1994.
 PF 20-APR-1994; U04387.
 PR 20-APR-1993; US-052449.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Daggett LP, Ellis SB, Liaw CM, Lu C;
 DR WPI: 94-341863/42.
 DR N-PSDB: Q79405.
 PT Isolated DNA encoding a human N-methyl-D-aspartate receptor
 PT subunit - used as probes in the identification and isolation of
 PT nucleic acids encoding related receptor subunits.
 PS Claim 11: Page 99-104; 156pp: English.
 CC The amino acid sequence of the novel N-methyl-D-aspartate (NMDA) receptor
 CC 2C (NMDAR2C) gene clone pcMV-26-scat-21-NotI-24. This clone covers the
 CC same sequence as the NMDAR2C sequence but contains a deletion of a.a.
 CC 722-8 of the NMDAR2C protein (R66035). A series of variants (Q79403-7)
 CC of the NMDAR2C receptor were constructed, by recombination, based on the
 CC sequence of the 4 basic clones: NMDA21 (Q79399), NMDA22 (Q79400), NMDA24
 CC (Q79401) and NMDA26 (Q79402). The NMDA receptor contains two subunits:
 CC subunit R1 (Q79370) and subunit R2 selected from the subunits 2A
 CC (Q79375), 2B (Q79377), 2C and 2D (Q79378). The receptor forms part of a
 CC family of NMDA receptors which have cation-selective channels and bind
 CC glutamate and NMDA. The expression of the genes allows the
 CC reconstruction of the NMDA receptor. The complete receptor can be used to
 CC identify compounds which bind or are antagonistic to the human NMDA
 CC receptor.
 CC Sequence 1219 AA;
 SO

Query Match 8.6%; Score 96; DB 12; Length 1219;
 Best Local Similarity 42.5%; Pred. No. 1.78e+01;
 Matches 17; Conservative 11; Mismatches 12; Indels 0; Gaps 0;

(TM)

Distribution rights by Oxford Molecular Ltd

generated

1 MCHSRCHPTMTILQAPTPA.....EPDYALDSTFLQQHPAAF 156

Gap 11

195121 seqs, 42852602 residues

Listing first 45 summaries

1:plr1 2:plr2 3:plr3 4:plr4

Mean 36.831; Variance 125.421; scale 0.294

lived by analysis of the total score distribution.

SUMMARIES

Category	Batch Length	DB	ID	Description	Pred. No.
9	99	7	156	2 UC5537	This is a DE line.
9	97	5	153	2 S53363	This is a DE line.
9	10	5	474	2 S51592	This is a DE line.
9	10	2	351	2 B34768	This is a DE line.
9	9	9	228	2 S55504	This is a DE line.
9	8	2	1729	2 M5TQ2Q	This is a DE line.
9	8	2	733	2 UC0032	This is a DE line.
9	9	2	913	2 S20450	This is a DE line.
9	9	2	1106	2 UC0045	This is a DE line.
9	9	2	1206	2 S24407	This is a DE line.
9	9	2	1468	2 S11515	This is a DE line.
9	9	1	268	2 B39429	This is a DE line.
9	9	1	417	2 E30341	This is a DE line.
9	9	1	816	2 F31277	This is a DE line.
9	9	0	333	2 S61849	This is a DE line.
9	9	0	439	2 S51939	This is a DE line.
9	9	0	1492	2 A40333	This is a DE line.
8	9	9	272	2 AA0021	This is a DE line.
8	8	8	272	2 S28013	This is a DE line.
8	8	351	2 S50754	This is a DE line.	
8	8	390	1 Q05E77	This is a DE line.	
8	8	475	2 A51151	This is a DE line.	
8	8	516	2 UC3332	This is a DE line.	

24	98	8.8	517	2	A55121	This is a DE line.	8.17e+00
25	98	8.8	975	2	B33131	This is a DE line.	8.17e+00
26	98	8.8	1486	2	B0333	This is a DE line.	8.17e+00
27	97	8.7	1239	1	Q0BE10	This is a DE line.	9.95e+00
28	97	8.7	1337	1	I38670	This is a DE line.	9.95e+00
29	97	8.7	1337	2	I52599	This is a DE line.	9.95e+00
30	96	8.6	245	1	MAW15	This is a DE line.	1.12e+00
31	96	8.6	245	1	MAW195	This is a DE line.	1.12e+00
32	96	8.6	256	2	E70032	This is a DE line.	1.12e+00
33	95	8.5	234	1	AUDH0	This is a DE line.	1.47e+00
34	95	8.5	316	2	G02424	This is a DE line.	1.47e+00
35	95	8.5	336	2	S55611	This is a DE line.	1.47e+00
36	95	8.5	377	2	A48018	This is a DE line.	1.47e+00
37	95	8.5	449	2	S16748	This is a DE line.	1.47e+00
38	95	8.5	493	2	JH0158	This is a DE line.	1.47e+00
39	95	8.5	534	2	S20961	This is a DE line.	1.47e+00
40	95	8.5	643	2	S52910	This is a DE line.	1.47e+00
41	95	8.5	3666	2	B48205	This is a DE line.	1.47e+00
42	95	8.5	3869	2	A48205	This is a DE line.	1.47e+00
43	94	8.4	232	2	JU0069	This is a DE line.	1.79e+00
44	94	8.4	389	2	S27200	This is a DE line.	1.79e+00
45	94	8.4	473	2	S50755	This is a DE line.	1.79e+00

ALIGNMENTS

RESULT	ID	SEQUENCE	STANDARD;	PRT;	156 AA.
1	JC5537				
XX	AC	xxxxxx			
XX	DT	01-JAN-1900			
XX	DE	This is a DE line.			
CC	A:	Experimental source: monocyte			
CC	A:	Note: the authors translated the codon CCG for residue 106 as Arg			
CC	SEQUENCE	156 AA; 16914 MW; 127043 CN;			

Query Match	Score	DB 2;	Length
99.78;	1111;	DB 2;	156;

```

Best Local Similarity 99.48; Pfile, NO. 0.10e-119;
Matches 155; Conservative 1; Mismatches 0; Indels 0; Gaps 0.

```

Db 1 MCHSRSCPTMTILQAPTAPASTIPGPRGSGPEIFTFDPLPEPAAPACAPRSGSGRHKK

1 MCHSRSCHTMTILQAPTPAPSTIPCBRRSGPELFTFDP LPEPAAAPAGRPASRGHRK 60

Db 61 RSRRLVPRVVRQLPVEEDNPAPAKRLIFLLITVFCQILMAEGVPAPLPEDADNAST 120

61 RSRRLVPRVVRKQLPVEEPNPAKRLLELLLTIVFCQILMAEGVPALDPEDAPNAASL 120

Db 121 APTVSPVLEPFNLTSEPSDYALDSTFLQHPAAE 156

QY 121 APTVSPVLEPENTSEPSDYALDSTFLQHPAAF 156

RESULT 2

ID	S33363	STANDARD;	PRT;	153	AA.
XX					

AC	XXXXXX
XX	
DE	01-TAN-1900

DI 01 JAN-1900
XX
DE This is a DE line.

XX A/Cross-references: EMBL:X67644

CC C;Genetics:
CC A;Introns: 70/3

CC C;Keywords: transmembrane protein
SQ SEQUENCE 153 AA; 16875 MW; 122478 CN;

Query Match	67.5%;	Score 752;	DB 2;	Length 153;
Next Local Classification:	70.6%;	Prod No. 1	100-74.	

BEST LOCAL SIMILARITY 10.0%; PREC. NO. 1.10E-14

[illegible]

```

xx    This is a DE line.
DE
xx
xx    TOIG of: b34768      check: 633       from: 1     to: 351
CC
CC
CC                                     >P1:B34768
CC   ORF5 protein - Orf virus (strain NZ)
CC   C.Species: Orf virus
CC   C.Date: 23-Aug-1991 #sequence_revision 23-Aug-1991 #text_change 31-Oct-1997
CC   C.Accession: B34768
CC   R.Frazer, K.M.; Hill, D.F.; Mercer, A.A.; Robinson, A.J.
CC   Virology 176, 379-389, 1990
CC   A>Title: Sequence analysis of the inverted terminal repetition in the genome o
CC   A.Reference number: A34768; MUID:90266454
CC   A.Accession: B34768
CC   A.Status: preliminary
CC   A:Molecule type: DNA
CC   A.Residues: 1-351 <FR>A
CC   A.Cross-references: GB:M30023; EMBL:M37623; NID:g332561; PID:g332564
SO   SEQUENCE          351 AA; 3988 MW; 427190 CN;

Query Match              10.2% Score 114; DB 2; Length 351;
Best Local Similarity 32.9%; Pred.No.3,09e+01;
Matches 24; Conservative 12; Mismatches 35; Indels 2; Gaps 2;

Db        236 QPRRAPPAAAGARCGRPQOQRORPVQAALARRAQROROPVRVRRRAR 295
Oy         15 QAPTPAPTIPDRGRSGEIFTFDPLPE-PAAPAGRPSASRGHRSRVLYPRV-VR 72
           ||| :|||
Db        296 ROGRAHORRGHR 308
Oy         73 ROLPYEEPNAKR 85

RESULT 5
XX ID S53504 STANDARD; PRT; 228 AA.
XX AC xxxxxx
XX DT 01-JAN-1900
XX XX
DE This is a DE line.
XX
XX TOIG of: s53504 check: 7615 from: 1 to: 228
CC
CC                                     >P1:S53504
CC   extensin-like protein S3 - alfalfa
CC   C.Species: Medicago sativa (alfalfa)
CC   C.Date: 15-Jul-1995 #sequence_revision 03-Nov-1995 #text-change 03-Nov-1995
CC   C.Accession: S53504
CC   R.Abraham, S.; Hayes, C.M.; Watson, J.M.
CC   Plant Mol. Biol. 27, 513-528, 1995
CC   A>Title: Expression patterns of three genes in the stem of lucerne (Medicago sa
CC   A.Reference number: S53504
CC   A.Accession: S53504
CC   A>Status: preliminary
CC   A:Molecule type: mRNA
CC   A.Residues: 1-228 <AB>
SO   SEQUENCE          228 AA; 22480 MW; 324273 CN;

Query Match             9.9% Score 110; DB 2; Length 228;
Best Local Similarity 24.0%; Pred.No.7,13e+01;
Matches 30; Conservative 36; Mismatches 56; Indels 3; Gaps 3;

Db        36 PPPTPATPPPTPOASPSPVOS-SBPYOSSPPYPQSSPPPAOSTPPPVQSSPPPVSA 94
Oy         16 APRPASPIPDRGSGEIFTFDPLPEAPAAPAGRPSASRGHKSRRLVLPVVRRQL 75
           ::::: |:::: |:::: |:::: |:::: |
Db        95 PVQOSPPTTPTTPPVYST-PPPASPPASPPSPSPPATPPATPPATPPATPPATPP 153
Oy         76 PYEPENRAKRLFLLTLITVFCCIIMAEGVFAFL-PPEDADANAASLAFTTPVSVDLEPNL 134
           ||::: |::: |::: |::: |::: |::: |::: |
Db        154 SSPPA 158
```

OY 135 TSEPS 139

RESULT 6
ID MSETGEQASNGEPODPTTIRVTKLTDREATVITIGLQDTIOSLIDGRREMIQSGFORVIAGRVL STANDARD

XX xxxxxx
DT 01-JAN-1900
DE This is a DE line.
XX

SO SEQUENCE 1729 AA; 194986 MW; 14801419 CN;

Query Match 9.8%; Score 109; DB 2; Length 1729;
Best Local Similarity 23.7%; Pred. No. 8.78e-01;
Matches 31; Conservative 30; Mismatches 65; Indels 5; Gaps 5;

DB 774 FLEKGRIPSTSSAPSTSENPPGSPFNS-EDPAIDIRAGRLPLGTRPNRTVRETVHFAAA 832

OY 13 ILQPTAPSTIPGPRGSGPEITFTDPLPEPAAPACR-PSASRGHKKRSRVLYPRV 71

DB 833 ARAESPNIISLTFTATHTFAPAGF-PLMASSNVSTSGPPGMPTRQVVSPTPTTGL 891

OY 72 RR-QLPVEEPNPAKRLFLTLTIVFCQILMAEGVPAPLP-PEDAPNNAASLAPTPVSPVL 129

DB 892 FEEDLSSGSSDQ 902

OY 130 EPNLTSEPSD 140

RESULT 7
ID J00532 STANDARD; PRT; 753 AA.

XX xxxxxx

DT 01-JAN-1900

DE This is a DE line.

XX
CC A:Accession: J00532
CC A:Molecule type: mRNA
CC A:Residues: 1-753 <DIN>
CC A:Experimental source: strain Jervis Bay isolate
SO SEQUENCE 753 AA; 82427 MW; 3187938 CN;

Query Match 9.2%; Score 102; DB 2; Length 753;
Best Local Similarity 32.9%; Pred. No. 3.67e+00;
Matches 24; Conservative 14; Mismatches 30; Indels 5; Gaps 5;

DB 271 LPRGNGRPGVLPGRKVGGA-PSS-NL-PLPTPOAPRRARERLQNSLHLHAGRONAPRLRP 327

OY 14 LQAPTAPSTIPGPR-RGSGPEITFTDPLPEPAAPACRPSASRGHKKRSRVLYPRV 72

DB 328 RRLR-SDPIDQTR 339

OY 73 ROLPVEEPNPAK 85

RESULT 8
ID S20590 STANDARD; PRT; 913 AA.

XX xxxxxx

DT 01-JAN-1900

DE This is a DE line.

XX TOIG of: s20590 check: 4303 from: 1 to: 913

CC >P1:s20590
CC exo-alpha-stallidase (EC 3.2.1.18) - Actinomyces viscosus

CC C:Species: Actinomyces viscosus

CC C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 09-Sep-1997

CC C:Accession: S20590

CC R:Henningsen, M.; Roggentin, P.; Schauer, R.

CC Biol. Chem. Hoppe-Seyler 372, 1065-1072, 1991

CC C:A:Title: Cloning, sequencing and expression of the stallidase gene from Actinon

CC C:A:Reference number: S20590; MUID:92162190

CC C:A:Accession: S20590

CC C:A:status: preliminary

CC C:A:Molecule type: DNA

CC C:A:Residues: 1-913 <HEI>

CC C:A:Cross-references: EMBL:X62276; NID:939254; PID:939255

CC C:Keywords: glycosidase; hydrolase

SO SEQUENCE 913 AA; 96216 MW; 3970532 CN;

Query Match 9.2%; Score 103; DB 2; Length 913;
Best Local Similarity 29.9%; Pred. No. 3.00e+00;
Matches 20; Conservative 14; Mismatches 31; Indels 2; Gaps 2;

DB 816 SPASRNAAPTPKGMPDEID-RPSDGTMAOPTGAP-ARRVDRRRRRPACGLARDQ 873

OY 16 APTTAPSTIPGPRGSGPEITFTDPLPEPAAPACRPSASRGHKKRSRVLYPRV 75

DB 874 RAADPGP 880

OY 76 PVEEPNP 82

RESULT 9
ID J00405 STANDARD; PRT; 1106 AA.

XX xxxxxx

DT 01-JAN-1900

DE This is a DE line.

XX
CC A:Accession: J00405
CC A:Molecule type: DNA
CC A:Residues: 1-1106 <SHI>
CC A:Cross-references: EMBL:X15867
CC C:A:Note: all the codons between two in-frame stop codons are translated; the tr

CC A:Note: the gene encoding this protein overlaps uvra gene

SO SEQUENCE 1106 AA; 119484 MW; 4705861 CN;

Query Match 9.2%; Score 102; DB 2; Length 1106;
Best Local Similarity 29.9%; Pred. No. 3.67e+00;
Matches 20; Conservative 16; Mismatches 29; Indels 2; Gaps 2;

DB 569 PGPAAGVAGRRPRVPOPRAGRHPLRRGPA-HPPGHTDRLRAGRPLRBAVHRPAPA 627

OY 19 PAPTTPGPRGSGPEITFTDPLPEPAAPACRPSASRGHKKRSRVLYPR-VVROLPV 77

DB 628 GQPPPHR 634

OY 78 EEPNPAK 84

RESULT 10
ID S24407 STANDARD; PRT; 1206 AA.

XX xxxxxx

DT 01-JAN-1900

DE This is a DE line.

XX TOIG of: s24407 check: 501 from: 1 to: 1206

CC >P1:S24407
CC formin isoform IV - mouse
CC C:Species: Mus musculus (house mouse)
CC C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 10-Sep-1997

```

Matches      22; Conservative      10; Mismatches      25; Indels      3; Gaps      3;

Db      142 CAQTVSLAAAAQPTPTSPPEPRGSGPPSLAP-GPVAEKGAAGKRGPPRGSPPEYQRER 200
       1 -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -| -|
Oy      7 CHPTWTIIQAAPT-PAPSTIIPGRGGS-GPEIFTFDPLLEPAALAPAGRPDSASGHKKRSR 64

RESULT      13
ID      E30341          STANDARD:          PRT;      417 AA.
AC      xxxxxx
XX
XX      01-JAN-1900
DE
DE      This is a DE line.
CC
CC      TOIG of: e30341 check: 5600 from: 1 to: 417
CC
CC      >P1,E30341
CC      alpha-1-adrenergic receptor - dog (fragment)
CC      C.Species: Canis lupus familiaris (dog)
CC      C.Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 08-Sep-1997
CC      C.Accession: E30341
CC      R.Libbert, F.; Parmentier, M.; Lefort, A.; Dinsart, C.; Van Sande, J.; Maenhaut
CC      Science 244, 568-572, 1989
CC      A.Title: Selective amplification and cloning of four new members of the G protein-
CC      A.Reference number: A30341; MUID: 89242119
CC      A.Accession: E30341
CC      A.Status: nucleic acid sequence not shown
CC      A.Molecule type: mRNA
CC      A.Residues: 1-417 <LI>
CC      A.Cross-References: GB:X14050; NID:g902; PID:g833784
CC      C.Superfamily: vertebrate rhodopsin
CC      C.Keywords: G protein-coupled receptor; glycoprotein; transmembrane protein
CC      SEQUENCE 417 AA; 45891 MW; 865241 CN;

Query Match      9.1%; Score 101; DB 2; Length 417;
Best Local Similarity 43.6%; Pred. No. 4,49e+00;
Matches 17; Conservative 10; Mismatches 10; Indels 2; Gaps 2;

Db      353 ALLSLAPAPPGRRG-RRDSCP-LTFRLALERGSPANG 389
Oy      12 TILQAPPAPTIPGRGSGPEIFTFDPLLEPAALAPAG 50

RESULT      14
ID      F31277          STANDARD:          PRT;      816 AA.
AC      xxxxxx
XX
XX      01-JAN-1900
DE
DE      This is a DE line.
CC
CC      TOIG of: f31277 check: 4471 from: 1 to: 816
CC
CC      >P1,f31277
CC      regulatory protein qa-1F - Neurospora crassa
CC      N.Alternate names: QA activator
CC      C.Species: Neurospora crassa
CC      C.Date: 26-Apr-1989 #sequence_revision 26-Apr-1989 #text_change 12-Sep-1997
CC      C.Accession: S04256; f31277
CC      R.Geevers, R.F.; Hulet, L.; Baum, J.A.; Tyler, B.M.; Patel, V.B.; Rutledge, B.J
CC      J Mol. Biol. 207, 15-34, 1989
CC      A>Title: DNA sequence, organization and regulation of the qa gene cluster of N
CC      A.Reference number: S04250; MUID: 89293848
CC      A.Molecule type: DNA
CC      A.Residues: 1-816 <GEZ>
CC      A.Cross-references: EMBL:X14603; NID:g3060; PID:g3068
CC      C.Genetics:
CC      A.Gene: qa-1F
CC      C.Superfamily: unassigned GAL4-type zinc cluster proteins; GAL4 zinc binuclear

```

MUSEUM
(TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

MSPrch_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 17 13:39:01 1998; MasPar time 12.21 Seconds

538.123 Million cell updates/sec

Tabular output not generated.

Title: >US-08-799-910-10

Description: (1-156) from US08799910.pep

Perfect Score: 1114

Sequence: 1 MCHSRCHPTMTILOAPRA.....EPSDYALDSTFLOOHRAAF 156

Scoring table: PAM 150

Gap 11

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

1:sp_fungi 2:sp_human 3:sp_invertebrate 4:sp_mammal
5:sp_mhc 6:sp_organelle 7:sp_phase 8:sp_plant
9:sp_bacteria 10:sp_protist 11:sp_virus 12:sp_vertebrate
13:sp_unclassified

Statistics: Mean 42.617; Variance 98.163; scale 0.434

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	1114	100.0	156	2	PROG1 PROTEIN (DIF-2 PR	8.46e-174
2	1114	100.0	351	11	HOMOLOGUE OF RETROVIRA	2.07e-02
3	1112	100.1	115	9	PREDICTED ORF.	3.65e-02
4	1112	100.1	285	1	CELL WALL PROTEIN PREC	3.65e-02
5	1111	100.0	3247	11	UL36.	4.84e-02
6	1110	99.9	225	11	GENOME, PARTIAL SEQUE	6.41e-02
7	1110	99.9	228	8	PROLINE RICH PROTEIN P	6.41e-02
8	108	9.7	452	8	MYROSINASE-BINDING P	1.12e-01
9	102	9.2	518	3	SIMILAR TO CUTICULAR C	5.74e-01
10	102	9.2	568	3	ARTICULIN P60.	5.74e-01
11	102	9.2	640	11	VACCINIA VIRUS GENE F1	5.74e-01
12	102	9.2	680	11	MC0181.	5.74e-01
13	103	9.2	913	9	SIALIDASE (EC 3.2.1.18	4.39e-01
14	103	9.2	1711	9	ENDOGALACTANASE A (EC 3.	4.39e-01
15	101	9.1	1426	10	CUT-LIKE 2 (CUX-2).	7.51e-01
16	100	9.0	185	8	PUTATIVE PROLINE-RICH	9.79e-01
17	100	9.0	333	8	HRPV.	9.79e-01
18	100	9.0	439	8	CHITINASE PRECURSOR.	9.79e-01
19	100	9.0	1048	19	CTD-BINDING SR-LIKE PR	9.79e-01
20	100	9.0	1491	12	ALPHA-1 TYPE II' COLLA	9.79e-01

ID	RESULT	1	PRELIMINARY:	PRT:	156 AA.
AC	092691;				
DT	01-FEB-1997 (TREMBLREL. 02, CREATED)				
DT	01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)				
DT	01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)				
DE	PROG1 PROTEIN (DIF-2 PROTEIN).				
GN	PROG1 OR DIF-2.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;				
OC	EUTHERIA; PRIMATES.				
RN	(1)				
RA	SEQUENCE FROM N.A.				
RA	TRAUZOLD A.;				
RL	SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.				
RP	SEQUENCE FROM N.A.				
RA	PIETZSCH A., BUECHLER C., ASLANIDIS C., SCHMITZ G.;				
RA	BIOCHEM. BIOPHYS. RES. COMMUN. 235:4-9(1997).				
RP	SEQUENCE FROM N.A.				
RA	KONDRATEV A.D., CHUNG K.N., JUNG M.O.;				
RA	CANCER RES. 56:1498-1502(1996).				
DR	EMBL: Y14511; E33102; -				
DR	EMBL: X96438; E350480; -				
SO	SEQUENCE 156 AA; 16928 MW; 46013510 CRC32;				

ALIGNMENTS

Query Match	100.0%	Score 1114:	DB 2:	Length 156:
Best Local Similarity 100.0%;				
Pred. No. 8.46e-174;				
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Db	1	MCHSRCHPTMTILOAPRASTIPGRSGEITFPDLEPAAPAGRSASGRHK 60		
Qy	1	MCHSRCHPTMTILOAPRASTIPGRSGEITFPDLEPAAPAGRSASGRHK 60		
Db	61	RSRRVLYPRVVRQQLVEEENPAKRLFLTLITVFCOIIAEEGVAPLPEDAPVASTL 120		
Qy	61	RSRRVLYPRVVRQQLVEEENPAKRLFLTLITVFCOIIAEEGVAPLPEDAPVASTL 120		
Db	121	APTPVSVLEPENTLSEPSDYALDSTFLOOHRAAF 156		
Qy	121	APTPVSVLEPENTLSEPSDYALDSTFLOOHRAAF 156		

```

RESULT 2
ID 085302 PRELIMINARY: PRT: 351 AA.
AC 085302:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE HOMOLOGUE OF RETROVIRAL PSEUDOROTEAASE.
OS ORF VIRUS.
OC VIRIDAE: DS-DNA ENVELOPED VIRUSES: POXVIRIDAE; CHORDOPOXVIRINAE;
OC PARAPOXVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX FRASER M., HILL D.F., MERCER A.A., ROBINSON A.T.;
RA MEDLINE: 90266454.
RL VIROLOG 176:379-389(1990).
DR EMBL: M30023; G332564;
SQ SEQUENCE 351 AA: 39888 MW: 8EEF741B CRC32;

Query Match 10.1%; Score 114; DB 11; Length 351;
Best Local Similarity 32.9%; Pred. No. 2.07e-02;
Matches 24; Conservative 12; Mismatches 35; Indels 2; Gaps 2;

Db 236 OPERARRAAGARGRGAPRQOQROPRVQRAAQAQROROPPRVARRARAR 295
15 QAPTPAPSTIPGRSGSEIFTFDLPPE-PAAPAGRESASGHRKRSRVLVPRV-VR 72
Db 296 RQORAHQRRGRGR 308
11 : : : : :
QY 73 RQLPVEEPNPKR 85

RESULT 3
ID P95306 PRELIMINARY: PRT: 115 AA.
AC P95306:
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DE 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)
DE PREDICTED ORF.
OS MYCOBACTERIUM TUBERCULOSIS.
OC PROKARYOTA: FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA TYAGI J.S.;
RL SUBMITTED (DEC-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA VERMA A., DASGUPTA N., AGGARWAL A.N., PANDE J.N., TYAGI J.S.;
RL INDIAN J. BIOCHEM. BIOPHYS. 32:429-436(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA TYAGI J.S.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: X63508; E282228;
SQ SEQUENCE 115 AA: 12106 MW: AF306566 CRC32;

Query Match 10.1%; Score 112; DB 9; Length 115;
Best Local Similarity 25.4%; Pred. No. 3.65e-02;
Matches 16; Conservative 23; Mismatches 22; Indels 2; Gaps 2;

Db 8 PGOALPPPPPPPPVPPNPAPPPLDSSPRLPTAPSPRRKPNPAPARRRRTAALR 67
11 : : : : :
QY 9 PTMTILOAPTAPSTIPGRSGSEIFTFDPE-LPEPAAPAGRPSA-SRGHRKRSRVL 66
11 : : : : :
Db 68 YRR 70
11 : : : : :
QY 67 YPR 69

RESULT 4
ID P78977 PRELIMINARY: PRT: 285 AA.

```

```

AC P78977:
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE CELL WALL PROTEIN PRECURSOR.
OS YAROMIA LIPOLYTICA (CANDIDA LIPOLYTICA).
OC EUKARYOTA: FUNGI: ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CX39-74A; TISSUE-PINA240;
RX MEDLINE: 97127825.
RA RAMON A., GIL R., BURGAL M., SENTANDREU R., VALENTIN E.;
RL YERST 12:1535-1548(1996).
DR EMBL: Z81006; E274837;
KM SIGNAL.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 285
SQ SEQUENCE 285 AA: 30036 MW: 9AFCA87 CRC32;

Query Match 10.1%; Score 112; DB 1; Length 285;
Best Local Similarity 28.5%; Pred. No. 3.65e-02;
Matches 43; Conservative 29; Mismatches 70; Indels 9; Gaps 9;

Db 122 ICHTKTVLSTEVITTPPVQTTTPAVKPTPEVVKPEPTPEV-PGVKPEPTGCP 180
11 : : : : :
QY 1 MCHSRSCPTMTILOAPTAP-STITPGRRSGSEIFTFDLPPEPAAPAGRPSASRGHR 59
11 : : : : :
Db 181 APKPEPEVVKPEPTPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEVPEV 238
11 : : : : :
QY 60 KRSRRVLYPRVYRROLP-VVEBNPAKRLILLITVFCOILMAEGVPA-DLPEDANA 117
11 : : : : :
Db 239 LSLPPSRSSPSLPLSPTLPPTPSRL-LS 268
11 : : : : :
QY 118 ASLAP-TP-VSPVLEPFNLTSEPSDYALDS 146

RESULT 5
ID Q65553 PRELIMINARY: PRT: 3247 AA.
AC Q65553:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE UL36.
OS BOVINE HERPESVIRUS TYPE 1.
OC VIRIDAE: DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COOPER;
RA SCHWYZER M., VLCEK C., LOWERY D.E., BELLO L.J., MEYER G., MISRA V.;
RL THIRY E., PACES V.;
RN [2]
RP SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 1-179 FROM N.A.
RC STRAIN-COOPER;
RA SCHWYZER M., STYGER D., VOGT B., LOWERY D.E., SINARD C.;
RA LABOISSIERE S., MISRA V., VLCEK C., PACES V.;
RL VER. MICROBIOL. 0:0-0(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-COOPER;
RA SCHWYZER M.;
RN [5]
RP SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [6]
RP SEQUENCE OF 1-179 FROM N.A.
RC STRAIN-JURA;
RA SCHWYZER M., STYGER D., VOGT B., LOWERY D.E., SINARD C.;
RA LABOISSIERE S., MISRA V., VLCEK C., PACES V.;
RL VER. MICROBIOL. 53:67-77(1996).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN-JURA;
RA SCHWYZER M., VLCEK C., LOWERY D.E., BELLO L.J., MEYER G., MISRA V.;

```



```

FT CHAIN      8      228      PROLINE RICH PROTEIN.
SQ SEQUENCE   228 AA;  22480 MW;  2A2B03B9 CRC32;.

Query Match          9.9%; Score 110; DB 8; Length 228;
Best Local Similarity 24.0%; Pred. No. 6,41e-02;
Matches           30; Conservative           36; Mismatches           56; Indels           3; Gaps           3;

Db    36 PPPTPANTPTTPOASPPPVOS-SPPVSSSPPPVSSSPPPAOSTPPPVSSPPVSAPP 94
      :|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY  16 APRTAPSTIGPRGSGPEITFEDPLPEPAAPAGRPSASRGHRSRRRLYPVVVRQL 75
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db    95 PVQOSSPPPTLTPPVOST-PPTASPPAPSPPPSPPTATPPATPPATPALTPPL 153
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY  76 PVEEPPNAKRLLELLTLTFVCQILMAEEGVAPL-PEDAPNAAISLAIPVSPVLPEPNL 134
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Db    154 SSPA 158
QY  135 TSERS 139

RESULT      8
ID 096343 PRELIMINARY; PRT; 552 AA.
AC 096343:
DT 01-FEB-1997 (TREMBLREL_02, CREATED)
DT 01-FEB-1997 (TREMBLREL_02, LAST SEQUENCE UPDATE)
DE MYOSESINASE-BINDING PROTEIN RELATED PROTEIN (FRAGMENT).
OS BRASSICA NAPUS (RAPE).
OC EUKARYOTA; PLANTA; EMPHYPHYTA; ANGIOSPERMAE; Dicotyledoneae;
OC CAPRALES; CRUCIFERAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-20516 OF SVALOFES KARAT;
RA TAIPALENSU J., FALK A., EK B., RASK L.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR ENBL, U59446; 1655830; -.
FT NBL TER 1
FT SEQUENCE 552 AA; 59270 MW; 6861EF25 CRC32;

Query Match          9.7%; Score 108; DB 8; Length 552;
Best Local Similarity 43.5%; Pred. No. 1.12e-01;
Matches           20; Conservative           11; Mismatches           12; Indels           3; Gaps           3;

Db    295 LRTPAPASPAPGPAPAPAGSHP-APAPAPAPAGGGPPAPAPG 339
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY  14 LQAVTPAPTIPGPRGSGPEITFEDPLPEPAAPA-G-RPSASRG 57

RESULT      9
ID 001662 PRELIMINARY; PRT; 418 AA.
AC 001662:
DT 01-JUL-1997 (TREMBLREL_04, CREATED)
DT 01-JUL-1997 (TREMBLREL_04, LAST SEQUENCE UPDATE)
DT 01-JUL-1997 (TREMBLREL_04, LAST ANNOTATION UPDATE)
DE SIMILAR TO CUTICULAR COLLAGEN.
DE 128F2.6.
OS CAENORHABDITIS ELYGANS.
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKS M.,
RA BONFIELD J., BURTON M., CONNELL M., CORSEY T., COOPER J.,
RA COULTON A., CRATON M., DEAR S., HAWKINS F., HILLIER L., JIER M.,
RA FULTON L., GARDNER A., GREEN P., HAMKINS F., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAU J., KIRSTEN J., LAISTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MCORMURY A., MORIMORE B.,
RA O'CALLAGHAN M., PARSONS J., PERCY C., RIFEEN L., ROOPRA A.,
RA SAUNDERS D., SHOWNKEEN R., SALLON N., SMITH A., SONNHAMMER E.,
RA STRADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M.,
RA VAGHANAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
```


Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES

Location/Qualifiers
1. .553
/organism="Homo sapiens"
/clone="260175"

source

MRNA
BASE COUNT 153 a 165 c 106 g 127 t 2 others

ORIGIN

Query Match 40.3%; Score 495; DB 12; Length 553;
Best Local Similarity 96.4%; Pred. No. 0.00e+00;
Matches 528; Conservative 0; Mismatches 17; Indels 3; Gaps 2;

Db 5 GACAGTAATCAATTTATTTGTTGTCACAGAACATACAGGAGATGTCAGTGCCTC 64
|||||
Cp 1204 GACAGTAATCAATTTATTTGTTGTCACAGAACATACAGGAGATGTCAGTGCCTC 1145
Db 65 CGTACAGAGCCACCAACCCCAACCCCTCTACCTCGCAGCCACCTTAAGGCGACTTCAAG 124
|||||
1144 CGTACAGAGCCACCAACCCCAACCCCTCTACCTCGCAGCCACCTTAAGGCGACTTCAAG 1085
125 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGATGCGCGAGATCTCAGCAGTGA 184
|||||
Cp 1084 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGATGCGCGAGATCTCAGCAGTGA 1025
Db 185 CAGACGAGATTGAGATCTGAGAGATCAGTACATCTCTAACTTACAGCCACACAG 244
|||||
Cp 1024 CAGACGAGATTGAGATCTGAGAGATCAGTACATCTCTAACTTACAGCCACACAG 965
Db 245 ACTTATCCCAAGCCGAGACTCTCTCCCAACCGAGTCTCCCATTTCTTCTCTACT 304
Cp 964 ACTTATCCCAAGCCGAGACTCTCTCCCAACCGAGTCTCCCATTTCTTCTCTACT 905
Db 305 TGGCGAGTCCAGGATCTCTGCTTCCACAGTCCCAAGGCTCAATTAATACAGAG 364
Cp 904 TGGCGAGTCCAGGATCTCTGCTTCCACAGTCCCAAGGCTCAATTAATACAGAG 845
Db 365 ACCTGATTTACAGCAGGAGGAGACATCTCAACCCCTGATTAAGTTAAATATATATAC 424
Cp 844 ACCTGATTTACAGCAGGAGGAGACATCTCAACCCCTGATTAAGTTAAATATATATAC 785
Db 425 GTACACATCTCCATCAGGAGAGAGTACATTAATTAATTAATTAAGGAGCAAT 484
|||||
Cp 784 GTACACATCTCCATCAGGAGAGAGTACATTAATTAATTAATTAAGGAGCAAT 725
Db 485 AAG--GAATTAATTAAGAGAGCTTTCCTTCCNACGGGGCTGAGCCAGCTGGAGT 542
|||||
724 AATAGAAATTAATTAAGAGAGCTTTCCTTCCNACGGGGCTGAGCCAGCTGGAGT 666
543 GCCTCGGT 550
|||||
Cp 665 GCCTCGGT 658

RESULT 2
LOCUS M60982 459 bp mRNA EST 07-JUN-1996
DEFINITION z698h09.s1 Pancreatic Islet Homo sapiens cDNA clone 339233 3'
ACCESSION M60982
NID g1367741
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 459)
Haller, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Matra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasaks, E., Waterston, R., Williamson, A., Woldmann, P. and Wilson, R.
TITLE Mashu-Merck EST Project
JOURNAL Unpublished (1995)

COMMENT

Contact: Wilson RK
Mashu-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: EMP1mer

FEATURES

High quality sequence stop: 226.
Location/Qualifiers
1. .459
/organism="Homo sapiens"
/note="Organ: pancreas; Vector: pBluescript SK-; Site-1: EcoRI; Site-2: XhoI; Reference: Hum Mol Gen 2, 1795 (1993)
Takeda et al. Cloned unidirectionally. Primer: Oligo dt.
-5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'
/db_xref="taxon:9606"
/clone="339233"
/tissue="pancreatic islet"
/lab_host="SOBR cells (kanamycin resistant)"
complement(<1..>459)

source

MRNA
BASE COUNT 121 a 147 c 82 g 108 t 1 others

ORIGIN

Query Match 36.6%; Score 450; DB 14; Length 459;
Best Local Similarity 98.9%; Pred. No. 0.00e+00;
Matches 454; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1 TTTTCTTTTGTGACAGTAATCAATTTATTTGTTGTCACAGAACATACAGGAGACT 60
|||||
Cp 1216 TTTTCTTTTGTGACAGTAATCAATTTATTTGTTGTCACAGAACATACAGGAGACT 1157
Db 61 CGACATCGCTCGGTACAGCCACCAACCCCAACCCCTCTACCTCGCAGCCACCTTAA 120
|||||
Cp 1156 CGACATCGCTCGGTACAGCCACCAACCCCAACCCCTCTACCTCGCAGCCACCTTAA 1097
Db 121 GGGGACTTCAAGAGATGAGAGATCTCAGGATCTCATTTGATGCGCGCGAAGTC 180
Cp 1156 GGGGACTTCAAGAGATGAGAGATCTCAGGATCTCATTTGATGCGCGCGAAGTC 1097
Db 121 GGGGACTTCAAGAGATGAGAGATCTCAGGATCTCATTTGATGCGCGCGAAGTC 180
Cp 1096 GGGGACTTCAAGAGATGAGAGATCTCAGGATCTCATTTGATGCGCGCGAAGTC 1037
Db 181 TCACAGTAGACAGAGAGAGTGAAGTCTGAGAGATCAGTACATCTCTAACTTACG 240
Cp 1036 TCACAGTAGACAGAGAGAGTGAAGTCTGAGAGATCAGTACATCTCTAACTTACG 977
Db 241 ACCCAGCCAGAGATCTTCCAGCCGAGAGTCTCTCCCAACCGAGTCTCCCATTT 300
|||||
Cp 976 ACCCAGCCAGAGATCTTCCAGCCGAGAGTCTCTCCCAACCGAGTCTCCCATTT 917
Db 301 CTTCCTCTACTTTGGCGAGTTCCAGTCTCTGCTTCCAGGATCTCCCAAGGCTCAAT 360
Cp 916 CTTCCTCTACTTTGGCGAGTTCCAGTCTCTGCTTCCAGGATCTCCCAAGGCTCAAT 857
Db 361 GAATACCAAGAGAGCTGATTTACAGCAGGAGGAGACATCTCAACCCCTGATTAAGTTAA 420
|||||
Cp 856 GAATACCAAGAGAGCTGATTTACAGCAGGAGGAGACATCTCAACCCCTGATTAAGTTAA 797
Db 421 AATACATATTACGTACATCTTCCATCTCCCTAGAGAGAC 459
|||||
Cp 796 AATACATATTACGTACATCTTCCATCTCCCTAGAGAGAC 758

RESULT 3
LOCUS AA047094 470 bp mRNA EST 06-SEP-1996
DEFINITION 2K74A02.r1 Soares pregnant uterus NBHPU Homo sapiens cDNA clone 488522 5'
ACCESSION AA047094
NID g1525011
KEYWORDS EST.

SOURCE	human.
ORGANISM	Homo sapiens
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;	
Vertebrata; Eutheria; Primates; Catarrhini; Homidae; Homo.	
REFERENCE	1 (bases 1 to 470)
AUTHORS	Hallier,L., Clark,N., Dubnue,T., Elliston,K., Hawkins,M., Holman,M., Holman,M., Kucaba,T., Le-M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Tevaskis,E., Waterston,R., Williamson,A., Woldmann,P. and Wilson,R. The Mashu-Merc EST Project Unpublished (1995)
TITLE	The Mashu-Merc EST Project
JOURNAL	
COMMENT	Contact: Wilson RK Mashu-Merc EST project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LINT ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Putative full length read The vector to vector length is 471 Seq primer: -28M13 revz from Amersham High quality sequence stop: 358. Location/Qualifiers 1..470 /organism="Homo sapiens" /note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5' AAGCGAGAATTCGCCGCGCCCTTTTTTTTTTTTTT 3'] , double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Palma Bonaldo." /db_xref="taxon:9606" /clone_id="488522" /clone_lib="Soares pregnant uterus NBDPU" /dev_stage="adult" /lab_host="DH10B"
MRNA	<1..>470
PAGE COUNT	103 a 82 c 141 g 137 t 7 others
IN	
Query Match	36.5%; Score 448; DB 14; Length 470;
Best Local Similarity	97.9% Pred. No. 0.00e+00;
Matches 456; Conservative	0; Mismatches 9; Indels 1; Gaps 1;
Dg	1 ATTATATGATTATTAATGAACCTCCTCAAGTGATGATGATGTGATGATTAATTTATTT 60
Oy	738 ATTTATATGATTATTAATGAACCTCCTCAAGTGATGATGATGTGATGATTAATTTATTT 797
Dd	61 TAACTATGACAAGGGTGTGAGATGTTCCCCTGCCTGATAATGACAGTCTCTGATTTTA 120
Oy	798 TAACCTATGACAAGGGTGTGAGATGTTCCCCTGCCTGATAATGACAGTCTCTGATTTTA 857
Dd	121 TTGACTTTGTGNACTGTGTGAGACANGACACCCTGNAACTGCGGCAAAGTAGAGANA 180
Oy	858 TTGACTTTGTGNACTGTGTGAGACANGACACCCTGNAACTGCGGCAAAGTAGAGANA 917
Dd	181 AATGGGAGGACTCGCGGTGNGGGAGAGAGCTCCCGGCTGGATGAAGTGTGGGCGGGGTC 240
Oy	918 AATGGGAGGACTCGCGGTGNGGGAGAGAGCTCCCGGCTGGATGAAGTGTGGGCGGGGTC 977
Dd	241 GTAAGTTAGAGAGTGCATNCATCCTCCAGCATCTCAACTCGCTGTCTACTGTGTAG 300
Oy	978 GTAAGTTAGAGAGTGCATNCATCCTCCAGCATCTCAACTCGCTGTGTCTACTGTGTAG 1037
Dd	301 ACTTGCGGGACCATTAAGAAATGAGATCCGATGATCCTTCCATCTTTGTAAGTCGCT 360

QY 1038 ACTTCGCGGACCATTTAGGAATGACATCCGTAGATCTTCCATCTTTTAAAGTCGCT 1097

Db 361 TTAGGGTGCGGCAGAGTAGAGGCTTAGGGCTTTGGTGCGCTCAGCAGACGACTGCG 420

OY 1098 TTAGAGTGCGCTGGAGAGTAGAGGTTGGGGGTT-GATGGCGCTGTACAGCAGCACTGCG 1156

Db 421 AGATCGCGCTAGTATGTTCTGTGTACACCAATAAATTGATTTACTG 466

OY 1157 AGATCGCGCTAGTATGTTCTGTGTGAMCAACAATAAATTGATTTACTG 1202

RESULT 4 AA457705 427 bp mRNA EST 06-JUN-1997

LOCUS Zx87f03.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 810724

DEFINITION 3.

ACCESSION AA457705

NID G2180425

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 427)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Stepiet,M., Tan,F., Theising,B., White,Y., Wyllie,T., Waterston,R. and Wilson,R.

TITLE Washu-Merck EST Project 1997

JOURNAL Unpublished (1997)

COMMENT Contact: Wilson RK
Washu-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAG Consortium (info@image.lln.gov) for further information.
Seq primer: -41m3 fwd. Et from Amersham
High quality sequence stop: 374.
Location/Qualifiers
1..427
organism="Homo sapiens"
note="Organ: ovary; Vector: p773D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5 TGTTCACATCTGAGAAGTGGGAGCGGCCGCGTTTTTTTATTTTATTTT 3']; double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p773 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo.
db_xref="taxon:9606"
clone="810724"
clone.lib="Soares ovary tumor NbHOT"
sex="female"
tissue_type="ovarian tumor"
lab_host="DH10B (ampicillin resistant)"
complement(<1..>427)
db_xref="GDB:6041404"

BASE COUNT 118 a 139 c 78 g 92 t

ORIGIN

Query Match 34.4%; Score 422; DB 23; Length 427;
Best Local Similarity 99.8%; Pied.No. 0.00e+00;
Matches 423; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 4 GACAGTAATCAATTTATTTGTTGTTCCAGAACATCTAGCGCATCTGCAGACTCGCTC 63

p 1204 GACAGTAATCAATTTATTTGTTGTTCCAGAACATCTAGCGCATCTGCAGACTCGCTC 1145

Db 64 CGTGACAGCCACCAACCCCAACCTCTACTGCGACGCCACCTTAAGGCGACTTAAG 123
 Cp 1144 CGTGACAGCCACCAACCCCAACCTCTACTGCGACGCCACCTTAAGGCGACTTAAG 1085
 Db 124 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGTCGCCGGAAGTCTCACAGTAGA 183
 Cp 1084 AAGATGGAAGATCTCAGGATCTCATTTCTTAATGTCGCCGGAAGTCTCACAGTAGA 1025
 Db 184 CAGACGAGATTGAGATCTGAGAGATGAGTCACTACCTCTTAATGAGCCACCAACG 243
 Cp 1024 CAGACGAGATTGAGATCTGAGAGATGAGTCACTACCTCTTAATGAGCCACCAACG 965
 Db 244 ACTTCATCCACAGCCGAGACGTCCTCCGCCACCGACGTCCTCCATTTCTTCTCTT 303
 Cp 964 ACTTCATCCACAGCCGAGACGTCCTCCGCCACCGACGTCCTCCATTTCTTCTCTT 905
 Db 304 TGCCGAGTTCCAGGTCCTCTGCTTCCACAGTCCCAAAAGCTCAATTAATACAGAG 363
 Cp 904 TGCCGAGTTCCAGGTCCTCTGCTTCCACAGTCCCAAAAGCTCAATTAATACAGAG 845
 Db 364 ACCTGATTACAGCAGGGGGAACATCTCAACCTTGATTAAGTTAAATTAATATTAC 423
 Cp 844 ACCTGATTACAGCAGGGGGAACATCTCAACCTTGATTAAGTTAAATTAATATTAC 785
 Db 424 GTAC 427
 Cp 784 GTAC 781
 RESULT 5
 LOCUS AA043722 460 bp mRNA EST 10-MAY-1997
 DEFINITION ZK59511.s1 Soares pregnant uterus NBHPU Homo sapiens cDNA clone
 ACCESSION AA043722
 NID 91521646
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Euthera; Primates; Catarrhini; Homnidae;
 Homo.
 REFERENCE 1 (bases 1 to 460)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,
 Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 Washu-Merck EST Project
 TITLE Unpublished (1995)
 JOURNAL
 COMMENT
 CONTACT: Wilson RK
 Washu-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1800
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1290 Std Error: 0.00
 Seq primer: -40M13 fwd: 381.
 High quality sequence stop: 381.
 Location/Qualifiers
 1. 460
 /organism="Homo sapiens"
 /note="Organ: uterus; Vector: p7773-Pac; Site: 1: Not I;
 Site: 2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo(dT) primer [5'-
 AACTGGAAGATTGCGGCGCGCTTTTCTTTTCTTTTCTTTT 3']
 AACTGGAAGATTGCGGCGCGCTTTTCTTTTCTTTTCTTTT 3']
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p7773 vector. Library
 went through one round of normalization. Library

constructed by M. Fatima Bonaldo."
 /db_xref="taxon:9606"
 /clone="487101"
 /clone_lib="Soares pregnant uterus NBHPU"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 complement(<1..>460)
 /db_xref="GDB:3760815"
 BASE COUNT 130 a 146 c 83 g 98 t 3 others
 ORIGIN
 Query Match 34.2%; Score 420; DB 19; Length 460;
 Best Local Similarity 98.3%; Pred. No. 0.00e+00;
 Matches 451; Conservative 0; Mismatches 3; Indels 5; Gaps 5;
 Db 2 ACAGTAATCAATTTTATTTGTTTCACAGAAATCTAGGCGATCTCAGACGCTCC 61
 Cp 1203 ACAGTAATCAATTTTATTTGTTTCACAGAAATCTAGGCGATCTCAGACGCTCC 1144
 Db 62 GTGACAGCCCAACCCCAACCCCTCTACCTGCGACGCCCTTAAGCGACTTCAGA 121
 Cp 1143 GTGACAGCCCAACCCCAACCCCTCTACCTGCGACGCCCTTAAGCGACTTCAGA 1084
 Db 122 AGATGAGAGATCTCAGGATCTCATTTCTTAATGAGTCCGCGAAGTCTCACAGTAGA 181
 Cp 1083 AGATGAGAGATCTCAGGATCTCATTTCTTAATGAGTCCGCGAAGTCTCACAGTAGA 1025
 Db 182 CAGACGAGTTGAGATCTGAGAGATGAGTCACTACCTCTTAATGAGCCACGACGAG 241
 Cp 1024 CAGACGAGTTGAGATCTGAGAGATGAGTCACTACCTCTTAATGAGCCACGACGAG 965
 Db 242 ACTTCATCCACAGCCGAGACGTCCTCCGCCACCGAGTCTCCCATTTCTTCTCTACT 301
 Cp 964 ACTTCATCCACAGCCGAGACGTCCTCCGCCACCGAGTCTCCCATTTCTTCTCTACT 905
 Db 302 TGCCGAGTTCCAGGTCCTCTGCTTCCACAGTCCCAAAAGCTCAATTAATACAGAG 361
 Cp 904 TGCCGAGTTCCAGGTCCTCTGCTTCCACAGTCCCAAAAGCTCAATTAATACAGAG 845
 Db 362 ACCTGATTACAGCAGGGGGAACATCTCAACCTTGATTAAGTTAAATTAATATT 421
 Cp 844 ACCTGATTACAGCAGGGGGAACATCTCAACCTTGATTAAGTTAAATTAATATT 787
 Db 422 ACCTGATTACAGCAGGGGGAACATCTCAACCTTGATTAAGTTAAATTAATATT 460
 Cp 786 ACCTGATTACAGCAGGGGGAACATCTCAACCTTGATTAAGTTAAATTAATATT 750
 RESULT 6
 LOCUS AA410666 463 bp mRNA EST 18-MAY-1997
 DEFINITION Z130909.r1 Soares ovary tumor NBHPU Homo sapiens cDNA clone 723904
 5' similar to SW:GL96_MOUSE P46694 IMMEDIATE EARLY PROTEIN GLY96.
 ACCESSION AA410666
 NID 92069789
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Euthera; Primates; Catarrhini; Homnidae;
 Homo.
 REFERENCE 1 (bases 1 to 463)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,
 Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
 White, Y., Wylie, T., Waterston, R. and Wilson, R.
 Washu-Merck EST Project 1997
 TITLE Unpublished (1997)
 JOURNAL
 COMMENT
 CONTACT: Wilson RK
 Washu-Merck EST Project
 Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -28ml3 rev2 ET from Amersham
 High quality sequence stop: 369.

FEATURES

source

1. 463
 /organism="Homo sapiens"
 /note="Organ: ovary; Vector: pT73D (Pharmacia) with a
 modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - Oligo(dt) primer [5']
 TGTACCAATCTGAGTGGAGCGCGCGGCTTTTCTTTTCTTTT
 TTTTACCAATCTGAGTGGAGCGCGCGGCTTTTCTTTTCTTTT
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library constructed by Bento Soares and
 M. Fatima Bonaldo."
 /db_xref="taxon:9606"
 /clone_lib="Soares ovary tumor NBH07"
 /sex="Female"
 /tissue_type="ovarian tumor"
 /lab_host="DH10B (ampicillin resistant)"
 <1. >463
 /db_xref="GDB:5935321"
 /db_xref="196 c 114 g 82 t
 BASE COUNT 71 a 196 c 114 g 82 t
 ORIGIN

Query Match 33.6%; Score 413; DB 22; Length 463;
 Best Local Similarity 97.4%; Pred. No. 0.00e+00;
 Matches 450; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

Db 6 ATGTGTACTCTCGACGTCCACCCGACATGACATCTCGACAGCCCGACCCGGCC 65
 1 ATGTGTACTCTCGACGTCCACCCGACATGACATCTCGACAGCCCGACCCGGCC 60
 Db 66 CCCGACCATCCCGGGGACCCCGGGGGT-CTCCGGTCTCGATGATCTTACCTTCGACC 124
 61 CCCGACCATCCCGGGGACCCCGGGGGTCTCCGGTCTCGATGATCTTACCTTCGACC 119
 Db 125 TCTCCGGAGCCCGCAGCGCCCTGTC-GGGCGCCCGACGCTCTCA-GCGGGACCGAAA 182
 120 TCTCCGGAGCCCGCAGCGCCCTGTC-GGGCGCCCGACGCTCTCGCGGGACCGAAA 179
 Db 183 GCGCA-CGCGAGGGTCTCTACCTCGAGTGTCCGGCGCAGTGCCTAGTCGAGAAC 241
 180 GCGAGCGCGAGGGTCTCTACCTCGAGTGTCCGGCGCAGTGCCTAGTCGAGAAC 239
 Db 242 GAACCCAGCCAAAGGCTTCTTCTGCTGCTCAACCATGCTTCTTTCGCGAGATCCGAT 301
 240 GAACCCAGCCAAAGGCTTCTTCTGCTGCTCAACCATGCTTCTTTCGCGAGATCCGAT 299
 Db 302 GCGTGAAGAGGGTGTCCGGCGCCCTCTCAAGAGAGCCCTTAAGCGCGATCCCT 361
 300 GCGTGAAGAGGGTGTCCGGCGCCCTCTCAAGAGAGCCCTTAAGCGCGATCCCT 359
 Db 362 GCGCGCCACCCCTGTGTCCCGTCTGAGAGCCCTTAATCTGACTTGGAGCCCTCGA 421
 360 GCGCGCCACCCCTGTGTCCCGTCTGAGAGCCCTTAATCTGACTTGGAGCCCTCGA 419
 Db 422 CTAGGCTTGAAGCTCAAGACTTCTCTCAACCAACACCCGGC 463
 420 CTAGGCTTGAAGCTCAAGACTTCTCTCAACCAACACCCGGC 461

RESULT 7 W52268 447 bp mRNA EST 11-OCT-1996
 LOCUS zc4610.r1 Soares senescent fibroblasts NBHSF Homo sapiens CDNA
 DEFINITION clone 325338 5' similar to SW:G196_MOUSE P46694 IMMEDIATE EARLY
 PROTEIN GLY96. [1] ;

ACCESSION

W52268
 91349380
 EST.

KEYWORDS

human.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 447)
 Hillier/L., Clark/N., Dubuque/T., Elliston/R., Hawkins/M.,
 Holman/M., Hultman/M., Kucaba/T., Le/M., Lennon/G., Marra/M.,
 Parsons/J., Rifkin/L., Rohlfing/T., Soares/M., Tan/F.,
 Treasaks/E., Waterston/R., Williamson/A., Woldmann/P. and
 Wilson/R.
 The WashU-Merck EST Project
 Unpublished (1995)

TITLE

WashU-Merck EST Project

JOURNAL

Unpublished (1995)

COMMENT

Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108.
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 3171 Std Error: 0.00
 Seq primer: mob.REGA+ET
 High quality sequence stop: 401.

FEATURES

source

1. 447
 /organism="Homo sapiens"
 /note="Vector: pT73D (Pharmacia) with a modified
 polylinker V.TYPE: phagemid. Site 1: Not I; Site 2: Eco
 RI; TGTACCAATCTGAGTGGAGCGCGGCTTTTCTTTTCTTTT
 3'] , double-stranded cDNA was size selected, ligated to
 Eco RI adapters (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of a modified pT73
 vector (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo."
 /db_xref="taxon:9606"
 /clone_lib="Soares senescent fibroblasts NBHSF"
 /tissue_type="senescent fibroblast"
 /lab_host="DH10B (ampicillin resistant)"
 <1. >447
 /db_xref="188 c 109 g 75 t 3 others
 BASE COUNT 72 a 188 c 109 g 75 t 3 others
 ORIGIN

Query Match 33.1%; Score 406; DB 15; Length 447;
 Best Local Similarity 96.7%; Pred. No. 0.00e+00;
 Matches 433; Conservative 0; Mismatches 12; Indels 3; Gaps 3;

Db 1 CTGCACTGCAACCCGACCATGACATCTCGACAGCCCGACCCCGCCCTTCACCA 60
 11 CTGCACTGCAACCCGACCATGACATCTCGACAGCCCGACCCCGCCCTTCACCA 70
 Db 61 TCCGGAGACCCCGCGCGCATCCGGTCTGATGATCTTACCTTCGACCTCTCCGAGC 120
 71 TCCGGAGACCCCGCGCGCATCCGGTCTGATGATCTTACCTTCGACCTCTCCGAGC 130
 Db 121 CCGAGAGGAGCCCATGCGCGGGCGCCCGAGGTCATGCGGGGAGCGAAGAGCGAGAG 180
 131 CCGAGAGGAGCCCATGCGCGGGCGCCCGAGGTCATGCGGGGAGCGAAGAGCGAGAG 189
 Db 181 AGGGTCTTACCTCTGAGTGTCCGGCGCAG-TGCCATCGAGAAACCGAACCGAC 239
 190 AGGGTCTTACCTCTGAGTGTCCGGCGCAGTGCAGTGCAGAACCGAACCGACCG 249
 Db 240 AAAAGGCTTCTTTTCTGCTGCTCAGCATGCTTCTTCCAGATCTGATGATGAGAG 299
 250 AAAAGGCTTCTTTTCTGCTGCTCAGCATGCTTCTTCCAGATCTGATGATGAGAG 309

Db	300	GGTGTGCGGGGCGCCCTCCGACAGAGAGCGCCCTTAAGCGCGATTCCTGGGCGCCAC	359
Oy	310	GGTGTGCGGGGCGCCCTCCGACAGAGAGCGCCCTTAAGCGCGATTCCTGGGCGCCAC	368
Db	360	CCCTGTGTCCCGCTCTCTCGAGAGCCCTTTAATCTGACTTCGGAGCCCTGGAGTACGCTCT	419
Oy	369	CCCTGTGTCCCGCTCTCTCGAGAGCCCTTTAATCTGACTTCGGAGCCCTGGAGTACGCTCT	428
Db	420	GGACCTCAGACACTTTCCTCCAGCAACAC	447
Oy	429	GGACCTCAGACACTTTCCTCCAGCAACAC	456
RESULT	8		
LOCUS	AA034911	481 bp * mRNA	EST 10-MAY-1997
DEFINITION	z425c01.91 Soares pregnant uterus NDHPU Homo sapiens cDNA clone		
ACCESSION	AA034911		
MID	g91506874		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;		
REFERENCE	1 (bases 1 to 481)		
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marie,M., Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.		
TITLE	Washu-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson RK Washu-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 1245 Std Error: 0.00 Seq primer: -40M13 fwd. from Amersham High quality sequence stop: 420. Location/Qualifiers 1. 481 /organism="Homo sapiens" /note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGGAAGAAATTCGCGCGCCCTTTTCTTTTCTTTTCTTTT 3'] , double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo." /db_xref="taxon:9606" /clone="471552" /clone_11b="Soares pregnant uterus NDHPU" /sex="female" /dev_stage="adult" /lab_host="DH10B" complement(<1. >481) /db_xref="GDB:3757554" BASE COUNT 140 a 139 c 85 g 105 t 12 others ORIGIN		
Query Match	32.7%	Score 402; DB 19; Length 481;	
Best Local Similarity	95.4%	Pred No: 0.00e+00;	
Matches 459; Conservative	0; Mismatches 13; Indels 9; Gaps 7;		
Db	1	GACAGTAATCAATTTTATTTGTGTTCACAGACATCTAGGCGATCTCGACAGTCGCTC	60

Cp	1204	GACAGTAAATCAATTGTTTGTGTTCACAGAACATACTAGCCGATCTCGAAGTCGC TC	1145
Db	61	CGTGACAGCCCCACCAACCACCCCTCTACTCGCAGCCACCCTTAAGGGACTTCAAG	120
Cp	1144	CGTGACAGCCCACCAACCACCCCTCTACTCGCAGCCACCCTTAAGGGACTTCAAG	1085
Db	121	AAGATGAAGAAGATCTCACGGATCTCATTTCTTAATAGTGTGCCGGAAGTCTCACACAGTGA	180
Cp	1084	AAGATGAAGAAGATCTCACGGATCTCATTTCTTAATAGTGTGCCGGAAGTCTCACACAGTGA	1025
Db	181	CAGACGGAGTTGAGATGCTGGAGGATGCAAGTCACCTCTTAAGTATGACACCACCACG	240
Cp	1024	CAGACGGAGTTGAGATGCTGGAGGATGCAAGTCACCTCTTAAGTATGACACCACCACG	965
Db	241	ACTTCNTCAACNCNCGANGAGGAGCGTCTCTCNCCANMGATGCTCTCCCATTTCTTCC	300
Cp	964	ACTTCATC---C-CAGCG-GGGAGAGTCTCTCCCAACCCAGAGTCTCTCTTCTTCC	910
Db	301	TACTTTGCCGNAGTTCCAGGTGCTGCTCTTCCACCACTCCACACAAGCTCAATTAATAC	360
Cp	909	TACTTTGCCGNAGTTCCAGGTGCTGCTCTTCCACCACTCCACACAAGCTCAATTAATAC	851
Db	361	CAAGAGACCTGGMATTTCACGACGAGGGGACATGTGAACCTTGGATAGTTAAAT	420
Cp	850	CAAGAGACCTG-CATTACACGACGAGGG-AAACATCTCA-CACCTTGCATAGTTAAAT	794
Db	421	AAATTTACGTACACATCTTCATCACCTCAGGAGGACGTACATAATACATAATATTN	480
Cp	793	AAATTTACGTACACATCTTCATCACCTCAGGAGGACGTACATAATACATAATATTN	734
Db	481	A 481	
Cp	733	A 733	
RESULT	9		
LOCUS	W52269	413 bp	mRNA EST 11-OCT-1996
DEFINITION	zc46a10.s1 Soares senescent fibroblasts NBHSF Homo sapiens CDNA clone 325338 3'.		
ACCESSION	W52269		
NID	G1349381		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 413)		
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,C., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasaki,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.		
TITLE	The Mashu-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson RK Mashu-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 3171 Std Error: 0.00 Seq primer: mob.REGA+ET. Location/Qualifiers 1..413 /organism="Homo sapiens" /note="vector: pRTT3D (Pharmacia) with a modified polylinker V-type; phagemid; Site_1: Not I; Site_2: Eco RI; TGtTACCAATCTGAAGGAGGCGCGCATTTTTTTTTTTTTTTTTT		
FEATURES	Source		

